

Unterrichtsplanungen von angehenden Lehrpersonen zum experimentellen Handeln

Eine videografiegestützte Analyse von Unterrichtsplanungen

Von der Pädagogischen Hochschule Heidelberg
zur Erlangung des Grades einer
Doktorin der Philosophie (Dr. phil.)
genehmigte Dissertation von

Josiane Tardent Kuster

aus

Zürich/Schweiz

2019

Erstgutachter: Prof. Dr. Markus Wilhelm

Zweitgutachter: Prof. Dr. Markus Rehm

Fach: Biologie und ihre Didaktik

Tag der mündlichen Prüfung: 12. November 2019

Abstract

Die Auseinandersetzung mit naturwissenschaftlichen Arbeits- und Denkweisen ist mit den neuen Bildungsstandards expliziter Bildungsauftrag für den naturwissenschaftlichen Unterricht auf der Sekundarstufe I. Das dafür notwendige Professionswissen soll in der Lehrpersonenausbildung aufgebaut werden und sich in Handlungssituationen wie Unterrichtsplanungen zeigen. Bisher hat sich bestätigt, dass die Analyse von schriftlichen Unterrichtsplanungen Aussagen über die Qualität von Unterrichtskonzeptionen ermöglichen. Jedoch ist wenig bekannt, über welches fachdidaktische Wissen angehende Lehrpersonen zur Planung von Unterricht zum Experimentieren im Sinne der Erkenntnisgewinnung verfügen und in welcher Form sich dieses, in Unterrichtsplanungen angewendete Wissen, erfassen lässt. In der vorliegenden Dissertation ist das im Projekt KUBeX¹ entwickelte, hoch inferente Ratinginstrument zur Beurteilung des entsprechenden fachdidaktischen Wissens faktoriell validiert und hinsichtlich der beiden, im Modell der «Didaktischen Rekonstruktion» angelegten Dimensionen «Analyse» und «Konstruktion» explorativ geprüft worden. Auch ist in den von 119 angehenden Lehrpersonen erstellten Unterrichtsplanungen und video-graphierten Planungsgesprächen die Qualität des erfassten fachdidaktischen Wissens untersucht worden. Die empirischen Befunde zeigen, dass sich auf der Grundlage der untersuchten Stichprobe das fachdidaktische Wissen zur Planung von Unterricht zum Experimentieren im Sinne der Erkenntnisgewinnung mit dem vorliegenden Instrument ermitteln lässt. Die im Modell der «Didaktischen Rekonstruktion» enthaltenen Dimensionen werden ausserdem durch die vorliegenden Ergebnisse gestützt. Auch geben die Ergebnisse zur Qualität des fachdidaktischen Wissens Anlass zur Vermutung, dass die Studierenden Schwierigkeiten bekunden, einen Unterricht zum Experimentieren im Sinne der Erkenntnisgewinnung zu planen oder dass sie dieses Wissen nicht explizit in den Planungen zeigen. Dabei ist auch ein Zusammenhang zwischen der Qualität ihres fachdidaktischen Wissens und den Vorgaben im Planungsraster festzustellen. Die eingangs erwähnten Zielsetzungen stellen vor diesem Hintergrund eine ernst zu nehmende Herausforderung für die Lehrpersonenausbildung und -weiterbildung dar.

¹ Kollegiales Unterrichtscoaching und Entwicklung experimenteller Kompetenz – Interventionsstudie in der schulpraktischen Ausbildung von Lehrpersonen für Biologie

Inhaltsverzeichnis

1	EINLEITUNG	9
TEILSTUDIE I		
2	THEORETISCHER UND EMPIRISCHER HINTERGRUND	17
2.1	NATURWISSENSCHAFTLICHE BILDUNG ALS BILDUNGS-AUFTRAG – ENTWICKLUNG VON BILDUNGSSTANDARDS	18
2.1.1	<i>Naturwissenschaftliche Grundbildung – das Konzept Scientific Literacy</i>	19
2.1.2	<i>Scientific Inquiry – Verankerung in den Bildungsstandards</i>	20
2.1.3	<i>HarmoS-Kompetenzmodell – Scientific Inquiry als nationaler Bildungsauftrag für den naturwissenschaftlichen Unterricht.....</i>	22
2.2	BILDUNGSSTANDARDS UND PROFESSIONALITÄT - BEDEUTUNG FÜR DIE NATUR- WISSENSCHAFTSDIDAKTISCHE FORSCHUNG UND LEHRPERSONENBILDUNG	25
2.2.1	<i>Professionalität und Professionalisierung von Lehrpersonen – eine begriffliche Annäherung.....</i>	25
2.2.2	<i>Professionalität im Lehrerberuf – unterschiedliche Ansätze mit Blick auf die naturwissenschaftsdidaktische Forschung</i>	26
2.3	PARADIGMEN DER EMPIRISCHEN PROFESSIONSFORSCHUNG – LEHRPERSONEN ALS ZENTRALE AKTEURE SCHULISCHER BILDUNGSPROZESSE	30
2.3.1	<i>Das Persönlichkeitsparadigma – Persönlichkeitsattribute der Lehrperson</i>	30
2.3.2	<i>Das Prozess-Produkt-Paradigma – das Lehrerhandeln und seine Wirkung</i>	31
2.3.3	<i>Das Expertenparadigma – Lehrpersonen als Expertinnen und Experten für das Lernen und Lehren.....</i>	32
2.4	PROFESSIONELLE HANDLUNGSKOMPENZ VON LEHRPERSONEN – MODELLIERUNGSANSÄTZE DER PROFESSIONSFORSCHUNG	34
2.4.1	<i>Berufsspezifische Dispositionen – eine Annäherung an den Kompetenzbegriff</i>	35
2.4.2	<i>Kompetenzmodellierungen in der Lehrpersonenbildungsforschung – Vorgehensweisen, Modelltypen und Mess-methoden.....</i>	39
2.4.3	<i>Professionelle Handlungskompetenz – Modellierungen mit Blick auf den Kompetenzbereich der Unterrichtsplanung</i>	43
2.5	PROFESSIONELLES WISSEN VON LEHRPERSONEN – KOGNITIVE DISPOSITION PROFESSIONELLER HANDLUNGSKOMPETENZ	51

2.5.1	<i>Taxonomie des Professionswissens von Lehrpersonen – zentrale Wissensdimensionen.....</i>	<i>52</i>
2.5.2	<i>Das fachdidaktische Wissen bzw. pedagogical content knowledge – Formen der Modellierung.....</i>	<i>53</i>
2.6	UNTERRICHTSQUALITÄT – EINFLUSS DES PROFESSIONSWISSENS VON LEHRPERSONEN.....	70
2.6.1	<i>Unterrichtsqualität – die Suche nach Qualitätsmerkmalen.....</i>	<i>71</i>
2.6.2	<i>Unterrichtsqualität – (fach-)didaktische Qualitätsmerkmale</i>	<i>72</i>
2.7	EXPERIMENTIEREN – EIN QUALITÄTSMERKMAL NATURWISSENSCHAFTLICHEN UNTERRICHTS.....	75
2.7.1	<i>Experimentelles Handeln – Scientific Inquiry als Fachinhalt</i>	<i>76</i>
2.7.2	<i>Experimentelle Kompetenzen – verschiedene Auffassungen und entsprechende Modellierungsansätze zu Scientific Inquiry.....</i>	<i>77</i>
2.8	UNTERRICHTSPLANUNGEN – EINBLICK IN DIE KOMPETENZEN DER STUDIERENDEN ...	81
2.8.1	<i>Unterrichtsplanung – ein antizipierendes Probehandeln für den Kompetenzaufbau von (angehenden) Lehrpersonen</i>	<i>81</i>
2.8.2	<i>Unterrichtsplanung – ein Ausdruck von Wissensbeständen in Form von Handlungsplänen oder Skripts.....</i>	<i>82</i>
2.8.3	<i>Unterrichtsplanungen – Vielfalt der Herangehensweisen bei der empirischen Untersuchung.....</i>	<i>85</i>
2.8.4	<i>Unterrichtsplanung – (fach-)didaktische Planungsmodelle als theoretischer Rahmen und Grundlage für die Planung</i>	<i>88</i>
2.8.5	<i>ERTE-Modell – ein fachdidaktisches Forschungsmodell zur Analyse von PCK (angehender) Lehrpersonen.....</i>	<i>95</i>
2.9	FAZIT	96
2.10	ÜBERBLICK ÜBER TEILSTUDIE I DER VORLIEGENDEN DISSERTATION	97
2.11	FORSCHUNGSFRAGEN	98
3	FORSCHUNGSDESIGN UND METHODEN	100
3.1	DAS PROJEKT KUBEX	100
3.2	TEILSTUDIE I – VALIDIERUNG EINES RATINGINSTRUMENTS ZUR ANALYSE VON UNTERRICHTSPLANUNGEN, -MATERIALIEN UND VIDEOGRAFIERTEN PLANUNGS- GESPRÄCHEN.....	102
3.2.1	<i>Zielsetzungen und Design.....</i>	<i>103</i>
3.2.2	<i>Stichprobe</i>	<i>104</i>
3.2.3	<i>Durchführung.....</i>	<i>105</i>
3.2.4	<i>Ratingmanual zur Erfassung des fachdidaktischen Wissens zur Planung von Unterricht zum experimentellen Handeln.....</i>	<i>107</i>
3.2.5	<i>Gütekriterien des Ratingmanuals</i>	<i>115</i>
3.2.6	<i>Statistische Verfahren</i>	<i>133</i>

4	ERGEBNISSE.....	137
4.1	VALIDIERUNG DES STRUKTURMODELLS ZUR ERFASSUNG FACHDIDAKTISCHEN WISSENS IN UNTERRICHTSPLANUNGEN, -MATERIALIEN UND VIDEOGRAFIERTEN PLANUNGSGESPRÄCHEN	138
4.1.1	<i>Datenanalyse – Voraussetzung für die faktorielle Validierung</i>	138
4.1.2	<i>Faktorielle Validierung des Ratingmanuals (1-Dimensionalität)</i>	139
4.1.3	<i>Faktorielle Validierung des Ratingmanuals (2-Dimensionalität)</i>	141
4.1.4	<i>1-Faktor- und 2-Faktorenmodell im Vergleich</i>	145
4.2	ZUSAMMENFASSUNG DER ERGEBNISSE DER TEILSTUDIE I	146
5	DISKUSSION.....	147
5.1	BEFUNDE ZUR FAKTORIELLEN VALIDIERUNG DES RATINGMANUALS.....	148
5.1.1	<i>Faktorielle Validierung des Manuals (Forschungsanliegen 1a).....</i>	149
5.1.2	<i>Faktorielle Validierung des Manuals (Forschungsanliegen 1b).....</i>	151
5.2	BEFUNDE ZUM DESIGN UND ZUR METHODISCHEN HERANGEHENSWEISE IN TEILSTUDIE I.....	155
5.3	GRENZEN UND AUSSAGEKRAFT DER TEILSTUDIE I.....	156
TEILSTUDIE II		
1	THEORETISCHER UND EMPIRISCHER HINTERGRUND	161
1.1	PROFESSIONSWISSEN – HANDLUNGSRESSOURCE VON (ANGEHENDEN) LEHRPERSONEN.....	162
1.1.1	<i>Das Fachwissen (content knowledge, CK).....</i>	163
1.1.2	<i>Das fachdidaktische Wissen (pedagogical content knowledge, PCK) ..</i>	166
1.2	PROFESSIONELLE KOMPETENZEN ZUM EXPERIMENTIEREN IM SINNE DER ERKENNTNISGEWINNUNG	168
1.3	DAS PLANUNGSHANDELN IM KONTEXT DER NATURWISSENSCHAFTSDIDAKTISCHEN LEHRPERSONENBILDUNG.....	171
1.4	FAZIT	175
1.5	ÜBERBLICK ÜBER TEILSTUDIE II	177
1.6	FORSCHUNGSFRAGEN UND HYPOTHESEN.....	178
2	FORSCHUNGSDESIGN UND METHODE	181
2.1	DESIGN	181
2.2	PROFESSIONSWISSENSTEST	183
2.3	STATISTISCHE VERFAHREN.....	184
3	ERGEBNISSE.....	186
3.1	VORKOMMEN UND QUALITÄT VON UNTERRICHTSPLANUNGEN ZUM EXPERIMEN- TELLEN HANDELN.....	187

3.1.1	<i>Fachdidaktisches Wissen zum experimentellen Handeln in Unterrichtsplanungen</i>	187
3.1.2	<i>Qualität fachdidaktischen Wissens zum experimentellen Handeln in Unterrichtsplanungen</i>	190
3.1.3	<i>Qualität in Unterrichtsplanungen – ein Vergleich von Interventions- und Kontrollgruppe</i>	194
3.2	ZUSAMMENHÄNGE ZWISCHEN DER PLANUNGSQUALITÄT UND DEN ERGEBNISSEN DER PROFESSIONSWISSENSTESTS	197
3.2.1	<i>Einfluss des deklarativen fachdidaktischen Wissens auf die Planungsqualität</i>	197
3.2.2	<i>Einfluss des deklarativen Fachwissens auf die Planungsqualität</i>	198
3.3	ZUSAMMENFASSUNG DER ERGEBNISSE DER TEILSTUDIE II	199
4	DISKUSSION	202
4.1	BEFUNDE ZUM VORKOMMEN UND ZUR QUALITÄT EXPERIMENTELLEN HANDELNS IN DEN PLANUNGSMATERIALIEN (FORSCHUNGSANLIEGEN 2A)	203
4.2	BEFUNDE ZU MÖGLICHEN ZUSAMMENHÄNGEN ZWISCHEN DER PLANUNGS- QUALITÄT UND DEN ERGEBNISSEN DER PROFESSIONSWISSENSTESTS (FORSCHUNGSANLIEGEN 2B)	205
4.3	GRENZEN UND AUSSAGEKRAFT DER TEILSTUDIE II	206
5	SYNOPSIS BEIDER TEILSTUDIEN UND AUSBLICK	209
5.1	SYNOPSIS DER BEIDEN TEILSTUDIEN I & II	210
5.1.1	<i>Zentrale Erkenntnisse der vorliegenden Studie (Teilstudie I & II)</i>	210
5.1.2	<i>Grenzen und Aussagekraft der vorliegenden Studie</i>	213
5.2	AUSBLICK UND MÖGLICHE IMPLIKATIONEN FÜR DIE LEHRPERSONENBILDUNG	215
ANHANG		221
A	TEILSTUDIE I	221
B	TEILSTUDIE II	227
ABKÜRZUNGSVERZEICHNIS		233
ABBILDUNGSVERZEICHNIS		235
TABELLENVERZEICHNIS		237
LITERATURVERZEICHNIS		239

1 Einleitung

Das Planen von Unterricht gehört zusammen mit seiner Durchführung und Auswertung zu den beruflichen Kernaufgaben und Tätigkeiten von Lehrpersonen (Baumert & Kunter, 2006; Blömeke, Felbrich & Müller, 2008; Greiten, 2015; Seel, 2011). Es wird als Schlüssel für erfolgreiches Unterrichten betont (Weingarten & van Ackeren, 2017), denn angemessenes Lehrerhandeln setzt immer auch professionelles Planungshandeln voraus (Wahl, 2002). Gut geplante Unterrichtsstunden können allerdings auch misslingen, genauso wie schlecht geplante Lektionen gelingen können. Mit einer durchdachten Unterrichtsplanung lassen sich jedoch viele der wichtigen Faktoren in den Blick nehmen, die dazu beitragen, die Wahrscheinlichkeit des Gelingens zu erhöhen (Zierer, Werner & Wernke, 2015a). Im Kontext der Lehrpersonenbildung und der Professionalisierungsforschung ist das Planungshandeln daher eine wichtige Schlüsselstelle für die Gestaltung von Unterricht (Zierer, Wernke & Werner, 2015b), die bislang allerdings in der Forschung nur unterschwellig Beachtung gefunden hat (Wernke & Zierer, 2017).

Das Planungshandeln, aufgefasst als eine Form didaktischen Denkens und Handelns (Wernke & Zierer, 2017), erfordert „die Ausdifferenzierung, Integration und Transformation“ des Professionswissens, insbesondere der drei Wissensdimensionen gemäß Shulman, (1987), zu denen das fachliche, das fachdidaktische und das pädagogische Professionswissen gezählt werden (Seel, 2011, S. 31). Dieses Professionswissen und darüberhinaus Faktoren wie Beliefs oder auch die arbeitsbezogene Motivation prägen den Erfolg von Lehrpersonen massgeblich (Kunter et al., 2013; Weinert, 1990). Daher sollten sich angehende Lehrpersonen die entsprechende Expertise, von der angenommen wird, dass sie erlernbar ist, im Rahmen ihrer Ausbildung aneignen können. Für die Entwicklung professioneller Handlungskompetenz und des damit verbundenen fachdidaktischen Wissens spielen Unterrichtsplanungen eine zentrale Rolle (Ball & Bass, 2000; Baylor, 2002). Im Kontext der Lehrpersonenbildung bieten sie eine geeignete und praxisnahe Möglichkeit, sich mit lernförderlichem Unterricht und den entsprechenden Qualitätskriterien auseinanderzusetzen. Ausserdem ermöglichen sie Einblicke in die kognitiven Aktivitäten und damit in die Planungskompetenzen von (angehenden) Lehrpersonen (König, Buchholtz & Dohmen, 2015; Tillema, 2009).

Das Planungshandeln erfolgt in der Lehrpersonenbildung meist auf der Basis überfachlicher, allgemeindidaktischer Modelle, die einen normativ festgelegten theoretischen Rahmen bilden (König, 2015; Zierer, 2012). Diese Modelle verbinden die Theorie mit der

Praxis. Sie reduzieren und strukturieren die Komplexität des zu planenden Unterrichts und bieten eine fundierte Grundlage zur reflektierten und theoriegeleiteten Planung und Analyse von Unterricht (Werner, Wernke & Zierer, 2017). Der Einsatz allgemeindidaktischer Modelle als Grundlage für die Gestaltung von Unterricht wird in der Literatur allerdings kontrovers diskutiert. In gewissen Untersuchungen werden die Modelle als unpraktikabel und von der Praxis als nicht relevant eingeschätzt (Bromme, 1981; Haas, 1998; Terbrügge, 2001), während neuere Untersuchungen zeigen, dass eine an allgemeindidaktischen Modellen orientierte Planung die Gestaltung eines qualitätsvollen Unterrichts massgeblich unterstützen kann (e.g. Zierer, Werner & Wernke, 2015a). In der Fachdidaktik der Naturwissenschaften wird das Modell der «Didaktischen Rekonstruktion», das auf der Basis allgemeindidaktischer Modelle entwickelt worden ist, aber den Fachinhalt stärker berücksichtigt, als Grundlage und Anleitung u.a. für die Planung von naturwissenschaftlichem Unterricht eingesetzt. Im Rahmen dieses Modells wird der wissenschaftliche Inhalt in Bezug auf zentrale Konzepte analysiert und unter Einbezug der Vermittlungsabsicht (Ziele) und der Schülervorstellungen in einen Kontext verpackt, der für die Schülerinnen und Schüler verständlich ist (Duit, Gropengiesser, Kattmann, Komorek, & Parchmann, 2012; van Dijk & Kattmann, 2007). Ausgehend von dieser «Analyse» erfolgt die didaktische Strukturierung bzw. «Konstruktion» konkreter und möglichst bedeutsamer Lerngelegenheiten, in denen lernförderliche Korrespondenzen und voraussehbare Lernschwierigkeiten sichtbar werden (Kattmann, Duit, Gropengiesser & Komorek, 1997, S. 12). Im Fokus des Modells der «Didaktischen Rekonstruktion» steht daher die Planung der Tiefenstruktur des Unterrichts, d.h. die Planung der Prozess- und Interaktionsebene zwischen Lehrenden und Lernenden (Kunter & Ewald, 2016; Kunter & Voss, 2011; Reusser & Halbheer, 2008). Das Modell basiert auf einem breiten theoretischen Fundament, das sich in der fachdidaktischen Forschung etabliert hat (Duit et al., 2012) und wird in der Ausbildung angehender Lehrpersonen als Planungsinstrument eingesetzt. Bislang ist aber weitgehend unklar, ob dieses Modell Einfluss auf die Qualität der Planungen und des Unterrichts hat (Stender, Brückmann & Neumann, 2015). Es wurde zudem empirisch nie geprüft.

Mit den neu zu erfüllenden Bildungsstandards (EDK, 2011; KMK, 2004, 2005) als Reaktion auf die schlechten Ergebnisse von PISA und TIMSS sind im naturwissenschaftlichen Unterricht in der Schweiz und in Deutschland die naturwissenschaftlichen Denk- und Arbeitsweisen (*Scientific Inquiry*) stärker in den Fokus des Unterrichtsgeschehens gerückt. Die Auseinandersetzung mit naturwissenschaftlichen Denk- und Arbeitsweisen, wie u.a. dem Experimentieren im Dienst der Erkenntnisgewinnung, erlangt damit im Naturwissenschaftsunterricht eine grössere Bedeutung. Für die Lehrpersonenbildung bedeutet dies, dass das dafür erforderliche fachdidaktische Wissen und auch die dafür notwendige Planungs-

kompetenz im Rahmen der Lehrpersonenbildung aufgebaut werden sollte. Für die Professionalisierungsforschung wird damit die Untersuchung, wie die entsprechenden professionellen Kompetenzen erworben und trainiert werden können, zur zentralen Aufgabe (Brunner et al., 2006, S. 76), darin eingeschlossen die Diagnose des Kenntnis- bzw. Kompetenzstandes angehender Lehrpersonen. Bislang ist die Untersuchung professioneller Handlungskompetenz vor allem durch die Analyse des Unterrichts (e.g. Börlin, 2012) oder auf der Basis von Tests (e.g. Tepner et al., 2012) erfolgt. Die Untersuchung der Planungskompetenz angehender Lehrpersonen stellt im Vergleich dazu ein noch völlig unbearbeitetes Feld der Kompetenzmessung und -modellierung dar (Aufschnaiter & Blömeke, 2010). Es ist wenig bekannt, über welches fachdidaktische Wissen zum Experimentieren im Sinne der Erkenntnisgewinnung angehende Lehrpersonen tatsächlich verfügen und in welcher Form sich dieses, in Unterrichtsplanungen angewendete Wissen, tatsächlich erfassen lässt. Die vorliegende Studie möchte diese Lücke schliessen und widmet sich vor diesem Hintergrund den im Projekt KUBeX von angehenden Lehrpersonen gestalteten Unterrichtsplanungen zum Experimentieren im Sinne der Erkenntnisgewinnung sowie den videografierten Planungsgesprächen über die erstellten Unterrichtsplanungen. Zwei Forschungsanliegen stehen dabei im Fokus der Untersuchung.

1. Validierung des Kompetenzstrukturmodells

- a. Das im Projekt KUBeX entwickelte Ratinginstrument zur Beurteilung von Unterrichtsplanungen, die das Ziel haben, Schülerinnen und Schüler Methoden der Erkenntnisgewinnung erwerben zu lassen, soll empirisch geprüft und soweit möglich validiert werden.
- b. Das Ratinginstrument soll weiter auch in Bezug auf die beiden zentralen Dimensionen des Modells der «Didaktischen Rekonstruktion» («Analyse», «Konstruktion») analysiert und empirisch geprüft werden.

2. Analyse von Unterrichtsplanungen

- a. Mit Hilfe des Ratinginstruments sollen erste Antworten auf die Frage gefunden werden, ob angehende Lehrpersonen Unterricht zum Experimentieren im Sinne der Erkenntnisgewinnung planen und wenn ja, welche Qualität fachdidaktischen Wissens sich in den Planungen zeigt.
- b. Diese Ergebnisse werden zudem auch in Bezug gesetzt zu weiteren, im Projekt KUBeX erfassten Faktoren, wie z.B. dem getesteten fachdidaktischen Wissen.

Die vorliegende Arbeit ist strukturell entlang der beschriebenen Forschungsanliegen in zwei Teilstudien gegliedert. Kapitelverweise sind jeweils mit den zu den Teilstudien passenden römischen Ziffern I bzw. II ergänzt. Der Aufbau der Arbeit ist in Abbildung 1 dargestellt.

In Teilstudie I wird das im Projekt KUBeX entwickelte Strukturmodell zur Erfassung fachdidaktischen Wissens zur Planung von Unterricht zum experimentellen Handeln auf der Grundlage der erhobenen Daten evaluiert und strukturell bzw. faktoriell validiert.

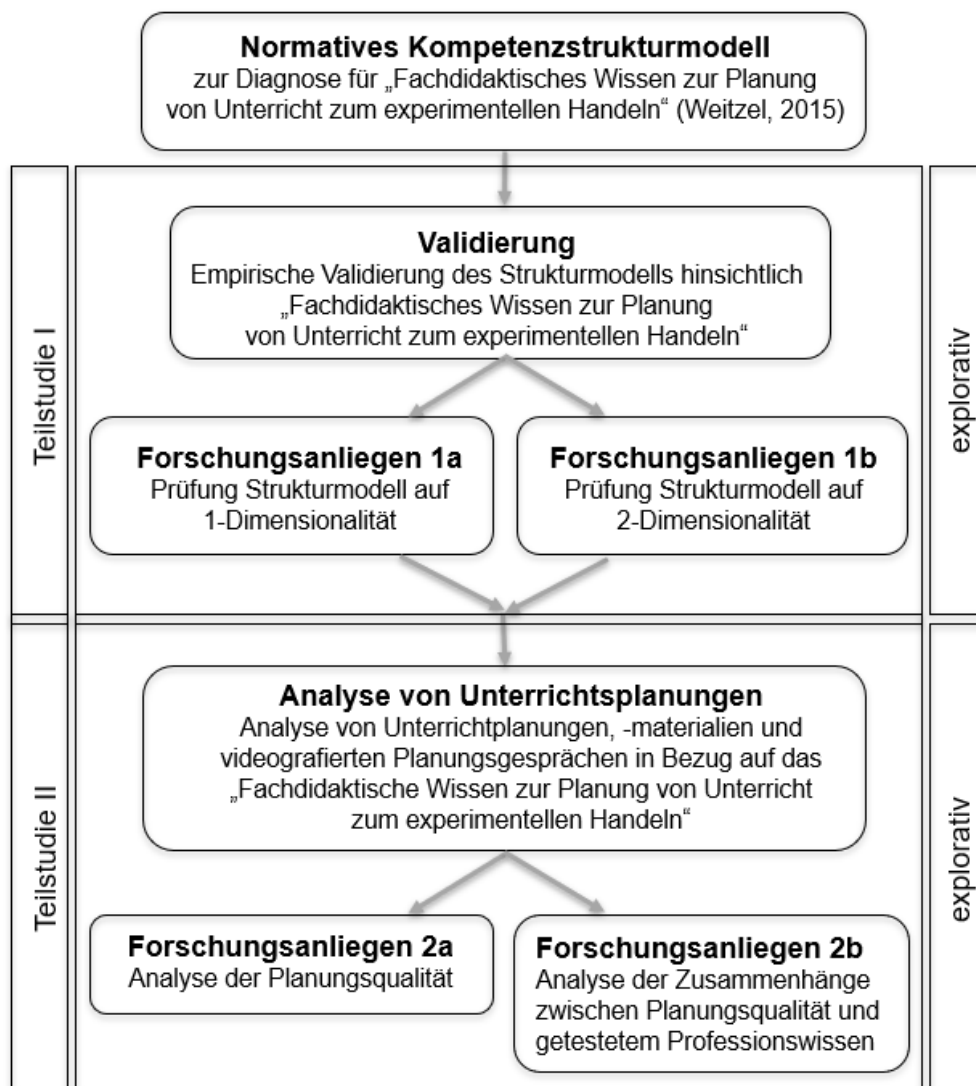


Abbildung 1: Chronologischer Aufbau der Arbeit mit den Forschungsanliegen der Teilstudien I & II

Mit dem Forschungsanliegen 1a wird explorativ geprüft, ob das modellierte Konstrukt das fachdidaktische Wissen zur Planung von Unterricht zum experimentellen Handeln tatsächlich als eine Dimension misst. Forschungsanliegen 1b prüft, ob sich in den erfassten Daten,

in Anlehnung an das Modell der «Didaktischen Rekonstruktion», zwei Subskalen bzw. Dimensionen, «*Analyse*» und «*Konstruktion*», extrahieren lassen.

Ausgehend von den in Teilstudie I gewonnenen empirischen Ergebnissen befasst sich Teilstudie II mit der Analyse von Unterrichtsplanungen, -materialien und videografierten Planungsgesprächen angehender Lehrpersonen. Untersucht wird dabei das in den Daten beobachtbare fachdidaktische Wissen zur Planung von Unterricht zum experimentellen Handeln. Forschungsanliegen 2a widmet sich dem Vorkommen und der Qualität der Planungsmaterialien. Mit dem Forschungsanliegen 2b werden mögliche Zusammenhänge zwischen der erfassten Planungsqualität und dem im Projekt KUBeX getesteten Professionswissen ermittelt.

Teilstudie I

2 Theoretischer und empirischer Hintergrund

Die Entwicklung einer naturwissenschaftlichen Grund- bzw. Allgemeinbildung (*Scientific Literacy*) im Sinne von *Science for all* wird schon seit längerem als wichtiges Bildungsziel und zentrale Aufgabe naturwissenschaftlichen Unterrichts erachtet (e.g. Fensham, 1986; Gräber & Nentwig, 2002; Labudde & Möller, 2012; Millar, 2006). Mittelmässige Ergebnisse internationaler Studien wie PISA und TIMSS haben die naturwissenschaftliche Bildung in den vergangenen Jahren verstärkt in den Fokus des Bildungsinteresses gerückt. Im englisch- wie auch im deutschsprachigen Raum sind daher in den vergangenen Jahren neue Bildungsstandards für den naturwissenschaftlichen Unterricht (EDK, 2011; KMK, 2004, 2005; National Research Council (NRC), 2012) entwickelt worden. Kern dieser verbindlichen Standards ist die Anwendung und Überprüfung naturwissenschaftlicher Konzepte und damit verbunden die kritische Reflexion der methodischen Vorgehensweisen und der Besonderheiten der Naturwissenschaften (Holbrook & Rannikmae, 2009; Labudde & Möller, 2012; Mayer, 2007). Das dafür notwendige Professionswissen und die damit verknüpfte professionelle Handlungskompetenz, insbesondere das Professionswissen zur Gestaltung entsprechender Lerngelegenheiten, sollen sich angehende Lehrpersonen während ihrer Ausbildung aneignen. Mit Unterrichtsplanungen besteht die Möglichkeit, Einblick in das Professionswissen und die Kompetenzen angehender Lehrpersonen zu erhalten (König et al., 2015; Tillema, 2009). Dazu bedarf es jedoch geeigneter Erhebungsinstrumente. Vor diesem Hintergrund befasst sich das Theoriekapitel mit den folgenden Themen:

Kapitel I, 2.1 thematisiert die neuen Bildungsstandards in Verbindung mit dem neuen Bildungsauftrag für den naturwissenschaftlichen Unterricht. Anschliessend wird ein Bezug zum Professionalitätsverständnis in der Lehrpersonenbildungsforschung hergestellt, insbesondere zum kompetenztheoretischen Professionalitätsverständnis der naturwissenschaftsdidaktischen Forschung (Kap. I, 2.2). Ausgehend davon werden drei Paradigmen der empirischen Professionsforschung genauer erläutert, welche die Lehrpersonen als zentrale Akteure schulischer Bildungsprozesse ansehen und beschreiben. Besonderes Augenmerk liegt auf dem Expertenparadigma, das aktuell in der Bildungsforschung, insbesondere der naturwissenschaftsdidaktischen Forschung, favorisiert wird (Kap. I, 2.3). In der Folge geht es um den Kompetenzbegriff in der Bildungsforschung und Modellierungsansätze professioneller Handlungskompetenz von Lehrpersonen (Kap. I, 2.4). Zur besseren Einordnung und Interpretation, aber auch zur Modellierung und Messung des Professionswissens, werden dann die drei zentralen Dimensionen des Professionswissens beleuchtet (Kap. I, 2.5). Dieses Wissen, insbesondere das fachdidaktische Wissen bzw. *pedagogical content*

knowledge (PCK), sollte sich in der Qualität des geplanten oder realisierten Unterrichts zeigen, d.h in der Performanz. Daher befasst sich das darauffolgende Kapitel zuerst mit der Unterrichtsqualität sowie allgemein- sowie fachdidaktischen Qualitätsmerkmalen, insbesondere dem Experimentieren (Kap. I, 2.6). Danach zeigt Kapitel I, 2.7 auf, wie experimentelle Kompetenz in der Literatur modelliert wird und davon ausgehend befasst sich Kapitel I, 2.8 mit Unterrichtsplanungen, einer handlungsnahen Möglichkeit, um Einblick in das themenspezifische Professionswissen der (angehenden) Lehrpersonen zu erhalten. Studien zum Thema Unterrichtsplanungen werden dabei im Hinblick auf die Erfassung fachdidaktischen Wissens bzw. von PCK beleuchtet und entsprechende Instrumente vorgestellt. In Kapitel I, 2.9 werden die zentralen Erkenntnisse als Grundlage für die vorliegende Studie zusammengefasst. Einen Überblick über die vorliegende Studie gibt Kapitel I, 2.10, in Kapitel I, 2.11 werden die konkreten Fragestellungen explizit ausformuliert.

2.1 Naturwissenschaftliche Bildung als Bildungsauftrag – Entwicklung von Bildungsstandards

Die Bedeutung naturwissenschaftlicher Bildung als Grundlage für eine effektive Teilnahme am gesellschaftlichen Leben wird schon seit vielen Jahren hervorgehoben (Bybee, 1997; Fischer, 1998; Holbrook & Rannikmae, 2009; Hurd, 1958; Lederman, Lederman & Antink, 2013; Millar, 1996). Grosse Veränderungen in unserer natürlichen Umwelt, die stark zunehmende Technisierung unseres Alltags, der Fachkräftemangel in Naturwissenschaften und Technik, wie auch das mittelmässige Abschneiden in internationalen Studien wie PISA und TIMSS, haben die naturwissenschaftliche Bildung in den vergangenen Jahren noch stärker in den Fokus des Bildungsinteresses gerückt (Labudde & Möller, 2012, S.12). Aus bildungspolitischer (Bybee, 1997; EDK, 2011; KMK, 2004, 2005; NRC, 2012) und naturwissenschaftsdidaktischer Sicht besteht weitgehend Einigkeit, dass die Entwicklung einer naturwissenschaftlichen Grund- bzw. Allgemeinbildung (*Scientific Literacy*) im Sinne von *Science for all* ein wichtiges Bildungsziel und eine zentrale Aufgabe des naturwissenschaftlichen Unterrichts darstellt (e.g. Fensham, 1986; Gräber & Nentwig, 2002; Labudde & Möller, 2012; Millar, 2006). Als Legitimierung hierfür ist primär die gesellschaftliche und kulturelle Bedeutung des Faches zu nennen. Entsprechend werden die lebensweltlichen Perspektiven der Lernenden zum Kristallisationspunkt naturwissenschaftlicher Bildungsprozesse (Dittmer, 2010, S. 13). Naturwissenschaftliche Grundbildung bietet darüber hinaus auch Perspektiven für eine spätere Berufswahl und ist gleichzeitig eine Grundlage für anschlussfähiges berufsbezogenes Lernen (Frank, Bernholt & Parchmann, 2016; Prenzel, 2000). Das erklärte Ziel von *Scientific Literacy* bleibt aber nach wie vor, „die Lebenswelt wissenschaftlich zu verstehen“ (Kattmann, 2003).

Die nächstfolgenden Abschnitte beleuchten das konzeptionelle Verständnis von *Scientific Literacy* und wie die neuen Bildungsstandards vor diesem Hintergrund normativ festgelegt und ausgearbeitet worden sind.

2.1.1 Naturwissenschaftliche Grundbildung – das Konzept *Scientific Literacy*

Wenngleich das Konzept *Scientific Literacy* in der Vergangenheit, aber auch heute noch unterschiedlich ausgelegt wird (DeBoer, 2000; Holbrook & Rannikmae, 2009), besteht in der *Scientific Community* dahingehend Konsens, dass naturwissenschaftliche Grundbildung verschiedene fachliche Kompetenzen und Wissensarten beinhaltet, inklusive überfachlicher Kompetenzen wie z. B. Problemlöse-, Reflexions- und Kooperationsfähigkeit. Auch affektive Komponenten wie Neugierde und Interesse an naturwissenschaftlichen Fragen gehören dazu (Labudde & Adamina, 2008; Labudde & Möller, 2012, S. 13). Das Modell von Bybee (1997) zu *Scientific Literacy* erweist sich als wichtige Grundlage für die aktuelle naturwissenschaftliche Bildung und Forschung. Das hierarchische Modell unterscheidet vier Stufen naturwissenschaftlicher Grundbildung (nominale, funktionale, konzeptionelle und prozedurale sowie mehrdimensionale *Scientific Literacy*), die keinen direkten Bezug zu konkreten Inhalts- oder Kompetenzbereichen aufweisen. Im Rahmen von PISA ist dieses Modell weiter ausdifferenziert worden. Seit PISA 2006 wird *Scientific Literacy* beschrieben als die Fähigkeit:

- naturwissenschaftliche Fragen formulieren zu können,
- sich neues Wissen aneignen und dieses anwenden zu können,
- naturwissenschaftliche Phänomene beschreiben und daraus Schlussfolgerungen ziehen zu können,
- charakteristische Eigenschaften der Naturwissenschaften als eine Form menschlichen Wissens und Forschens verstehen zu können,
- den Einfluss und die Bedeutung der Naturwissenschaften und der Technik in unserer materiellen, intellektuellen und kulturellen Umwelt erkennen zu können,
- der Bereitschaft, sich mit naturwissenschaftlichen Ideen und Themen zu beschäftigen und sich reflektierend mit ihnen auseinandersetzen zu können (Prenzel et al., 2007).

Die Beschreibung von *Scientific Literacy* verdeutlicht, dass primär nicht die Vermittlung von bloßem Faktenwissen und die Kenntnis von naturwissenschaftlichen Bezeichnungen und Begriffen im Zentrum stehen. Der tatsächliche Kern naturwissenschaftlicher Bildung ist einerseits die Anwendung und Überprüfung naturwissenschaftlicher Konzepte, auch im Hinblick auf deren Relevanz für den persönlichen Alltag und die Gesellschaft. Andererseits gehört die damit verbundene kritische Reflexion der methodischen Vorgehensweisen sowie der Besonderheiten der Naturwissenschaften dazu (Bodzin & Beerer, 2003; Bybee,

2002; Millar, 2006; OECD, 2007). Konzepte wie *Nature of Science* (Natur der Naturwissenschaften) und *Scientific Inquiry* (Erkenntnisgewinnung), aber auch wissenschaftliche Arbeitstechniken (*practical work*) sind somit zentrale Bestandteile naturwissenschaftlicher Grundbildung (Holbrook & Rannikmae, 2009; Labudde & Möller, 2012; Mayer, 2007). Die Erforschung und Untersuchung naturwissenschaftlicher Phänomene (*Scientific Inquiry*) und die damit verbundenen Erkenntnisprozesse erhalten eine grössere Bedeutung für den Unterricht (e.g. Lederman et al., 2013). Dies widerspiegelt sich auch in den neuen Bildungsstandards Deutschlands und der Schweiz.

2.1.2 *Scientific Inquiry* – Verankerung in den Bildungsstandards

Prozessbezogene und damit fachmethodische Fähigkeiten und Fertigkeiten haben im naturwissenschaftlichen Unterricht Deutschlands wie auch der Schweiz lange Zeit eine eher geringere Rolle gespielt und sind meist nur implizit mit den Fachinhalten vermittelt worden. Eine Folge davon waren mittelmässige Ergebnisse aus Schulleistungstudien wie TIMSS 1995 und PISA 2000, 2006. Diese Befunde haben international und auch national bildungspolitische Reformbewegungen ausgelöst und zu einer Abkehr von inputorientierten Lehrplänen hin zur Einführung von outputorientierten Bildungsstandards (EDK, 2011; KMK, 2004, 2005; National Board for Professional Teaching Standards [NBPTS], 2002) geführt. Mit den normativ festgelegten Bildungsstandards sind neue Anforderungen an das Lehren und Lernen in der Schule in Form von Kompetenzen festgelegt worden (cf. Klieme et al., 2003, S. 19). Diese Anforderungen sind präskriptiver Natur und folgen einem Bildungskonzept wie dem Konzept der naturwissenschaftlichen Grundbildung sowie stellenweise auch einer Fachlogik wie z.B. den Basiskonzepten (Mayer & Wellnitz, 2014). Dabei fokussieren sie auf das vom kompetenten Anwender gezeigte Verhalten und nicht auf die ihm zugrundeliegenden Fertigkeiten und Wissens Elemente (Frey, 2014, S. 714). Standards können auf einem theoretischen Kompetenzmodell basieren, werden in der Regel aber über die Sammlung, Bewertung und Strukturierung durch Experten und Stakeholder entwickelt (ebd.). Sie unterliegen gewissen A-priori-Annahmen, die bei Lehrplanzielen nicht oder weniger ausgeprägt sind (Gut-Glanzmann, 2012, S. 14). Es sind dies (a) die Lehr- und Lernbarkeit, (b) die Verbindlichkeit, (c) der Kompetenzbezug im Sinne einer Kompetenzausprägung, (d) die Messbarkeit, aber auch (e) die Skalierbarkeit von Standards (ebd.). Der Ausdruck Bildungsstandards wird damit in doppelter Weise gebraucht, als Ziel und als Mass der Zielerreichung (Oelkers & Reusser, 2008). Bildungsstandards beschreiben mit Hilfe von Kompetenzen oder Aufgabenstellungen Bildungsziele für die pädagogische Arbeit, ausgedrückt als erwünschte Lernergebnisse der Schülerinnen und Schüler im Sinne eines Performanzmasses. Sie konkretisieren damit den gesellschaftlichen Bildungsauftrag,

den allgemeinbildende Schulen zu leisten haben (Klieme et al., 2003, S. 19). Ziel dieser Outputorientierung ist letztendlich auch, Schülerleistungen aufgrund dieser festgesetzten Standards testbasiert überprüfen zu können (cf. Fend, 2011, S. 8).

Für den naturwissenschaftlichen Unterricht in Deutschland sind Bildungsstandards als Regelstandards für den mittleren Schulabschluss in Biologie, Chemie und Physik definiert und eingeführt worden (für die Biologie, KMK, 2005, cf. Tab. 1).

Tabelle 1: Kompetenzbereiche des Faches Biologie in Deutschland (KMK, 2005, S. 7)

Kompetenzbereiche	
Fachwissen	Lebewesen, biologische Phänomene, Begriffe, Prinzipien, Fakten kennen und den Basiskonzepten zuordnen
Erkenntnisgewinnung	Beobachten, Vergleichen, Experimentieren, Modelle nutzen und Arbeitstechniken anwenden
Kommunikation	Informationen sach- und fachbezogen erschliessen und austauschen
Bewertung	Biologische Sachverhalte in verschiedenen Kontexten erkennen und bewerten

Sie enthalten vier Kompetenzbereiche, wobei die Erkenntnisgewinnung (*Scientific Inquiry*) als Handlungsdimension in einem eigenen Kompetenzbereich verankert und mit entsprechenden Kompetenzen ausdifferenziert ist. Für den Naturwissenschaftsunterricht in der Schweiz wurden demgegenüber Grundkompetenzen für einen integrierten naturwissenschaftlichen Unterricht festgelegt und beschrieben (EDK, 2011). Es handelt sich dabei um Mindeststandards. Darin enthalten sind ebenfalls entsprechende Fähigkeiten und Fertigkeiten der Erkenntnisgewinnung. Diese wurden auf der Basis des HarmoS-Kompetenzmodells entwickelt und für das Integrationsfach Naturwissenschaften, insbesondere auch für die Biologie, als verbindliche und anzustrebende Standards ausformuliert und festgelegt.

Die Ausgangslage der Bildungsreform in der Schweiz, das HarmoS-Kompetenzmodell und die auf dieser Grundlage entstandenen Grundkompetenzen sowie der damit verbundene Bildungsauftrag für die Schulen und Lehrpersonenausbildung werden im nächsten Abschnitt ganz allgemein, aber auch mit Blick auf Kompetenzen zum Prozess der Erkenntnisgewinnung (*Scientific Inquiry*) näher dargelegt.

2.1.3 HarmoS-Kompetenzmodell – *Scientific Inquiry* als nationaler Bildungsauftrag für den naturwissenschaftlichen Unterricht

Auslöser für die Bildungsreform in der Schweiz war der Entscheid der Schweizerischen Konferenz der kantonalen Erziehungsdirektoren (EDK), die kantonalen Bildungssysteme mit sprachregionalen Lehrplänen, verbindlichen Standards und einem nationalen Bildungsmonitoring zu harmonisieren (Gut-Glanzmann, 2012; Ramseier, Labudde, & Adamina, 2011). Durch die Annahme des revidierten Bildungsartikels in der Bundesverfassung 2006 durch das Schweizer Stimmvolk und mit der Inkraftsetzung des HarmoS-Schulkonkordats 2009 wurden diese Reformen besiegelt (Gut-Glanzmann, 2012). Im Rahmen dieses gross angelegten bildungspolitischen Reformprojekts wurde 2008 das Kompetenzmodell HarmoS Naturwissenschaften⁺ durch das Konsortium HarmoS Naturwissenschaften⁺ entwickelt (HarmoS, 2008). In diesem Modell spiegeln sich die schulische Realität wie auch internationale Konzepte von *Scientific Literacy* (cf. Kap. I, 2.1.1) gleichermassen wider (Labudde & Adamina, 2008, S. 351). Auch orientiert sich das Modell an den Facetten des Kompetenzbegriffs von Weinert (2001; cf. Kap. I, 2.4.1). Das Modell ist in Abbildung 2 gezeigt und soll in der Folge näher erläutert werden.

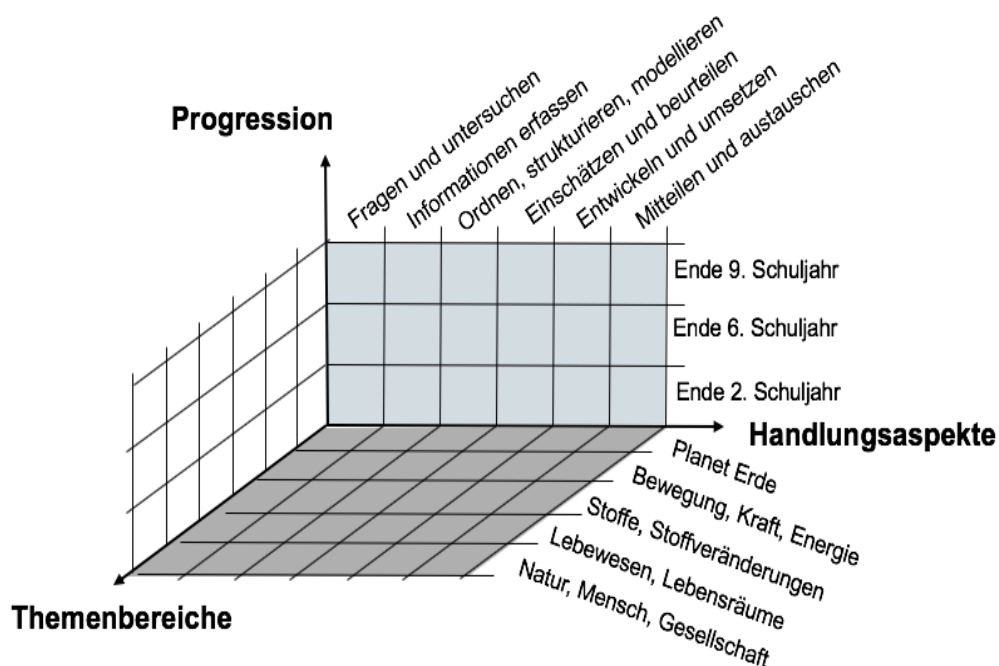


Abbildung 2: HarmoS Kompetenzmodell (Gut, 2012, S. 108, modifiziert durch die Autorin)

Das HarmoS-Kompetenzmodell besteht aus einer dreidimensionalen Matrix. Diese umfasst in der ersten Dimension die sogenannten Handlungsaspekte, in der zweiten Dimension die Themenbereiche und in der dritten Dimension die Progression. Die Dimension der

Handlungsaspekte beinhaltet insgesamt sechs spezifische Fähigkeiten und Fertigkeiten, die den Konzepten *Scientific Inquiry* und *Nature of Science* zuzuordnen sind. Diese werden verstanden als zentrale Fähigkeiten und Fertigkeiten im Sinne von „Werkzeugen“ des Denkens, Handelns und Fühlens, und beschreiben in ihrer Gesamtheit grundlegende Fähigkeiten und Fertigkeiten einer naturwissenschaftlichen Grundbildung. Die zweite Dimension der naturwissenschaftlichen Themenbereiche weist zentrale Bezugspunkte zu den verschiedenen inhaltlichen Bereichen der Naturwissenschaften auf. Darin eingeschlossen sind auch interdisziplinäre gesellschaftliche Anliegen wie Gesundheit, Technik und Technikfolgen sowie Umwelt und Nachhaltigkeit, deshalb die Bezeichnung Naturwissenschaften⁺. Die dritte Dimension, die sogenannte Progression, bezieht sich auf die erste Dimension des Modells, die Handlungsaspekte. Sie beschreibt Anforderungen an die Ausprägung von Kompetenzen (Performanz) in Form von Niveaus. Auf der Grundlage dieses Modells wurden Standards entwickelt, die nach einer breit geführten Vernehmlassung überarbeitet und als verbindliche Grundkompetenzen für die Naturwissenschaften 2011 von der EDK in Kraft gesetzt worden sind (EDK, 2011). Diese Grundkompetenzen wurden anschliessend in den Deutschschweizer Lehrplan (D-EDK, 2014) eingearbeitet, der im Oktober 2014 an die Kantone zur Einführung freigegeben wurde (Reusser, 2014). Mit der Einführung des kompetenzorientierten Lehrplans (D-EDK, 2014) und den zu erfüllenden Bildungsstandards (EDK, 2011) erlangte die Auseinandersetzung mit naturwissenschaftlichen Arbeits- und Denkweisen, wie u.a. dem Experimentieren im Dienst der Erkenntnisgewinnung, auch in der Schweiz eine zentrale Bedeutung. Die dafür anzustrebenden Kompetenzen sind im Handlungsaspekt „Fragen und untersuchen“ beschrieben und durch mehrere Teilaspekte festgelegt. Es sind dies die Teilaspekte:

- Bewusst wahrnehmen (Lebewesen, Gegenstände, Situationen, Prozesse betrachten, beobachten, beschreiben u.a.) und dabei Fragen, Probleme, Hypothesen aufwerfen.
- Erkundungen, Untersuchungen oder Experimente durchführen: Fragen und Probleme aufgrund von Beobachtungen und Vorkenntnissen aufwerfen; Erkundungen, Untersuchungen und Experimente planen und durchführen; Daten sammeln, messen, ordnen und auswerten, Hypothesen überprüfen bzw. Sachverhalte und Regelmäßigkeiten erkennen und festhalten, Folgerungen ziehen.
- Geeignete Werkzeuge, Instrumente und Materialien auswählen und verwenden: Für Erkundungen, Untersuchungen, Experimente und technische Konstruktionen.
- Ergebnisse zusammenstellen und über Ergebnisse und Untersuchungsmethoden nachdenken: Ergebnisse und Schlussfolgerungen aus Untersuchungen zusammentragen, Erkundungen und Experimente beurteilen und bewerten, Frage- und Problemstellungen, Versuchsanlagen, Untersuchungs- und Messmethoden sowie technische Konstruktionen reflektieren, hinterfragen und dazu Verbesserungen vorschlagen.

Die hier aufgeführten Teilaspekte gelten insbesondere auch für das Unterrichtsfach Biologie und sind in den Grundkompetenzen (EDK, 2011) für drei Zyklen (Ende des 4., 8. und 11. Schuljahres) weiter ausdifferenziert.

Ergebnisse aus internationalen Vergleichsstudien wie PISA 2006, 2012 und 2015 haben gezeigt, dass sich die naturwissenschaftlichen Leistungen von Schweizer Jugendlichen seit 2000 verbessert haben und dass sie signifikant über dem Durchschnitt der OECD-Länder liegen. Die Ergebnisse von 2015 sind allerdings aufgrund des Wechsels in der Erhebungsmethode (Papier-Tests zu Computer-Tests) und Unklarheiten bezüglich der Repräsentativität der Stichprobe gemäss der Schweizerischen Konferenz der kantonalen Erziehungsdirektoren (EDK) und dem Staatssekretariat für Bildung, Forschung und Innovation (SBFI) zu wenig gesichert. Grundsätzlich zeigt sich aber, dass die Leistungen der Jugendlichen in den Naturwissenschaften eher im Mittelfeld und hinter den erbrachten Leistungen in Mathematik anzusiedeln sind. Eine Expertise, die im Auftrag der Bildungsdirektion des Kantons Zürich wichtige Aspekte des Unterrichts im naturwissenschaftlichen und technischen Bereich beleuchten und Handlungsempfehlungen erarbeiten sollte, widerspiegelt ein ähnliches Bild (Stern, Zeyer & Metzger, 2009). Laut dieser Expertise wird das übergeordnete Ziel der Fächer im Bereich Natur und Technik oft nicht erreicht, nämlich die Schülerinnen und Schüler in die Lage zu versetzen, Vorgänge in der belebten und unbelebten Welt mit Hilfe grundlegender naturwissenschaftlicher Konzepte zu erklären und Ereignisse vorherzusagen. Darin eingeschlossen sind auch die Vertrautheit der Schülerinnen und Schüler mit naturwissenschaftlichen Denk- und Arbeitsweisen (*Scientific Inquiry*) sowie ein Bewusstsein für Möglichkeiten und Grenzen naturwissenschaftlicher Erkenntnisse (*Nature of Science*). Diese Befunde decken sich mit Ergebnissen aus anderen Studien, die zeigen, dass die Schülerinnen und Schüler in den Naturwissenschaften an Aufgaben scheitern, die ein Verständnis zentraler Konzepte und grundlegender Besonderheiten naturwissenschaftlichen Arbeitens verlangen (Hammann, 2004; Hammann, Phan, Ehmer & Grimm, 2008; Mayer, Grube & Möller, 2008). Es stellt sich vor diesem Hintergrund die Frage, über welche Kompetenzen Lehrpersonen verfügen müssen, um den Bildungsauftrag im Sinne des Handlungsaspekts „Fragen und untersuchen“ erfolgreich umzusetzen und welche Schlussfolgerungen sich daraus für die Ausbildung angehender Lehrpersonen ergeben. Die empirische Bildungsforschung liefert belastbare Hinweise, wonach Handlungen und Entscheidungen der Lehrperson die individuellen Entwicklungen und Übergänge im Bildungssystem erheblich und nachhaltig beeinflussen (Zlatkin-Troitschanskaia, Beck, Detlef, Reinhold & Mulder, 2009, S. 13). Um die beruflichen Anforderungen ganz allgemein, aber auch im Rahmen der Bildungsstandards erfüllen zu können, ist Professionalität im Handeln gefragt, insbesondere auch für die Förderung von Kompetenzen im Bereich der Erkenntnis-

gewinnung in den Naturwissenschaften bzw. im Fach Biologie. Das nächste Kapitel I, 2.2 widmet sich vor diesem Hintergrund den Themen Professionalität und Professionalisierung von Lehrpersonen. Es setzt sich mit den Fragen auseinander, welches Professionsverständnis der naturwissenschaftsdidaktischen Forschung zugrunde liegt und was in diesem Zusammenhang professionelles Handeln kennzeichnet.

2.2 Bildungsstandards und Professionalität - Bedeutung für die naturwissenschaftsdidaktische Forschung und Lehrpersonenbildung

Professionalität im pädagogischen Handeln scheint sich als zentrale Determinante für die Qualität von Bildungsprozessen und Bildungsergebnissen zu erweisen. Sie ist von einem komplexen Zusammenspiel von Kompetenzfacetten geprägt (Leuders, Nückles, Mikelskis-Seifert & Philipp, 2019). Die Frage, worin der Kern erfolgreichen beruflichen Handelns von Lehrpersonen besteht und was die pädagogische Professionalität in diesem Zusammenhang kennzeichnet, ist in den letzten Jahren kontrovers diskutiert worden (Helsper, 2007; Järvinen, 2014). Die nächstfolgenden Abschnitte widmen sich diesem Diskurs mit dem Ziel, die naturwissenschaftsdidaktische Forschung in diesem Spannungsfeld einzuordnen.

2.2.1 Professionalität und Professionalisierung von Lehrpersonen – eine begriffliche Annäherung

Professionalität und Professionalisierung sind Begriffe, die im Kontext der Beruflichkeit eng miteinander verbunden sind und unterschiedliche Prozesse und Zustandsbeschreibungen betreffen (Helsper, 2007; Järvinen, 2014). Die pädagogische Professionalität im Lehrerberuf wird als einen erreichten oder zu erreichenden Zustand „der Könnerschaft“ (Combe, 2006a, S. 37) beschrieben oder auch verstanden als „gekonnte Beruflichkeit“ mit einer „schwer bestimmbaren Schnittmenge ... aus Wissen und Können“ (Nittel, 2004, S. 351). „Könnerschaft [ist dabei] jedoch niemals blosser Wissensapplikation“, sondern verlangt auch „die Kontextualisierung dieses Wissens auf besondere Fälle“ (Neuweg, 2005, S. 206) – eine Fähigkeit, die neben Wissen auch Erfahrung, Reflexion und Persönlichkeit erfordert (Fischler, 2008). Cramer (2012, S. 23) subsumiert unter Professionalität die Fähigkeiten und Fertigkeiten, den Berufsalltag zu bewältigen und erachtet diese als Zielkriterium der Lehrpersonenausbildung im Sinne hoher professioneller (berufsbezogener) Kompetenz. Während die Professionalität von Lehrpersonen einen zu erreichenden oder erreichten Zustand umschreibt, verweist der Terminus Professionalisierung auf eine

Handlungsdynamik (Combe, 2006a) oder wie Nittel ausführt, den „Prozess in der Zeit“ (Nittel, 2004, S. 347).

Traditionell ist mit Professionalisierung der historische Weg zunehmender Akademisierung der Lehrpersonenbildung gemeint (Cramer, 2012, S. 23). Neuerdings wird der Terminus auch für die Beschreibung individueller Professionalisierungsprozesse verwendet und damit mit der professionellen Entwicklung der Lehrperson innerhalb der institutionalisierten Lehrpersonenbildung in Verbindung gebracht (ebd.). Die Professionalisierung kann daher auch als Ausbildung und Entwicklung einer besonderen Handlungsqualität im Sinne von Professionalität verstanden werden (Herking, 2015). Die Analyse der Faktoren, die Lehrpersonen in die Lage versetzen, die komplexen Aufgaben im beruflichen Alltag erfolgreich zu bewältigen, ist hier wegleitend. Der nächste Abschnitt geht auf dieses Thema ein.

2.2.2 Professionalität im Lehrerberuf – unterschiedliche Ansätze mit Blick auf die naturwissenschaftsdidaktische Forschung

Zur Beschreibung des professionellen Lehrerhandelns werden in der deutschen erziehungswissenschaftlichen Literatur verschiedene Ansätze und Theorien unterschieden und kontrovers diskutiert. Zwei auf den ersten Blick gegensätzliche Theriefamilien pädagogischer Professionalität und die damit verbundenen Forschungsmethoden erweisen sich als bestimmend für den Diskurs (Cramer, 2012, S. 24). Es sind dies die qualitative Forschung mit soziologischem Theoriehintergrund sowie die quantitative Forschung mit psychologischem Theoriehintergrund (Tillmann, 2011). Der ersten Theriefamilie wird u.a. der *strukturtheoretische Bestimmungsansatz* zugeordnet. Zur zweiten Theriefamilie, dem *kompetenztheoretischen Bestimmungsansatz*, gehören Ansätze, die sich am Persönlichkeits- bzw. dem Expertenparadigma ausrichten (cf. Combe & Paseka, 2012). Beide hier hervorgehobenen Ansätze sollen vor- und einander gegenübergestellt werden.

I. Das strukturtheoretische Verständnis von Professionalität

Das *strukturtheoretische* Verständnis pädagogischer Professionalität und die damit verbundene Forschungsrichtung basiert auf der soziologischen Theorie zum pädagogisch-professionellen Handeln von Oevermann (1996). Darin wird die pädagogische Praxis als Spezialfall der *sozialisatorischen Interaktion* im Sinne einer quasi-therapeutischen Tätigkeit und prophylaktischen Handlung aufgefasst (Oevermann, 1996, S. 148f.). Professionelles Handeln entwickelt sich im Wesentlichen aus den Aufgaben und Schwierigkeiten, die sich in der Struktur und Dynamik der pädagogischen Interaktion ergeben (Tenorth, 2013, S. 23). Das professionelle Handeln wird damit als Rekonstruktion der sich gegenseitig bedingen-

den Handlungsstruktur zwischen Lehrpersonen und Schülerinnen und Schülern gefasst (Helsper, 2011). Kerngeschäft der Lehrperson ist neben der Erziehung die „Wissens- und Normenvermittlung“ (Oevermann, 1996), wobei das Primat der Wissensvermittlung zukommt. Hierzu zählt Oevermann die Vermittlung von Erfahrungswissen, Traditionswissen, Kulturtechniken, Praktiken etc. (Oevermann, 1996, S. 144). Dieses Kerngeschäft wird auch als „Relationierung von Person und Sache (cf. Helsper & Hummrich, 2008) oder als eine Erfahrungsbewegung zwischen Ich und Gegenstand (Combe, 2006b, S. 34)“ beschrieben (zitiert in Helsper, 2007, S. 567f.). Individualdiagnostische Fragestellungen, gepaart mit einer interpretativ-rekonstruktiven methodischen Herangehensweise, sind hier wegleitend (Bölsterli et al., 2010, S. 277). Das Handeln der Lehrperson lässt sich nach diesem Verständnis nicht standardisieren, sondern zeigt seine Besonderheit in der Arbeit am konkreten Beispiel oder Fall (Vogelsang, 2014, S. 34). Der Erwerb pädagogischer Expertise erfolgt somit nicht durch routiniertes Handeln, sondern durch Reflexion (Järvinen, 2014, S. 10). Leitend dafür ist ein „forschender Habitus“ im Sinne einer „Habitualisierung eines neugierigen, skeptischen, forschend-fragenden Umgangs mit der Praxis“ (Helsper & Kolbe, 2002, S. 395). Terhart (2011, S. 207) unterstreicht dies mit der Aussage, dass die selbstkritische, reflektierende Rückwendung auf das eigene Handeln, auf die eigene berufliche Entwicklung ein zentraler Motor für die Weiter(!)-Entwicklung professioneller Fähigkeiten ist.

Die hier beschriebenen beruflichen Aufgaben und Anforderungen von Lehrpersonen, aber auch die „Generalisierung der Erziehungserwartung gegenüber Lehrern und Schule“ stehen im Widerspruch zu den Bedingungen der Pflichtschule (Baumert & Kunter, 2013; cf. Terhart, 2011, für eine genauere Explikation der Widersprüchlichkeit) und erzeugen vor diesem Hintergrund theoretische Dilemmata. Dies u.a. auch weil die Gelingensbedingungen alltäglichen Unterrichts damit nicht hinreichend erklärbar sind (Tenorth, 2006, 2013). Der *kompetenzorientierte* Professionalisierungsansatz setzt hierzu einen Kontrapunkt und begegnet diesem Widerspruch, indem er sich nach den Aufgaben der Pflichtschule richtet und auf empirischen Studien beruht, die zeigen, dass u.a. Fachwissenschaft und Fachdidaktik für die Qualität von Unterricht bedeutsam sind (e.g. Baumert & Kunter, 2006; Bromme, 2008; Lipowsky, 2006). Dieser forschungsmethodische Ansatz wird im nächsten Abschnitt näher erläutert.

II. Das kompetenztheoretische Verständnis von Professionalität

Im Fokus des *kompetenztheoretischen* Professionalisierungsansatzes steht das Aufgabenprofil des Lehrerberufs, das mit den institutionellen Vorentscheidungen über Ziele, Fächerstruktur, Stundentafel, Lehrplan u.a. eine spezifische und sachliche Rollenbeziehung zwischen der Lehrperson und den Schülerinnen und Schülern festlegt (Baumert & Kunter,

2006, S. 472). Kompetenzbereiche und Wissensformen, die zur Bewältigung der Aufgaben im Lehrerberuf als bedeutsam und notwendig erachtet werden, werden definiert (Terhart, 2011, S. 207). Dabei stellen in Anlehnung an Bromme (1997) die Vorbereitung, Inszenierung und Durchführung von Unterricht das Kerngeschäft der Lehrperson dar (Baumert & Kunter, 2006; Blömeke, 2002; Tenorth, 2006), wobei die Erziehungsleistung im Sinne des erziehenden Unterrichts nach Hebart (1806) darin enthalten ist (Baumert & Kunter, 2006). Im Zentrum dieses Ansatzes steht die Qualität von Unterricht und die Ergründung des professionellen Wissens erfolgreich handelnder Lehrpersonen (Kamm & Bieri, 2008).

Terhart schreibt dazu (Terhart, 2011, S. 207):

Professionell ist ein Lehrer dann, wenn er in den verschiedenen Anforderungsbereichen (Unterrichten und Erziehen, Diagnostizieren, Beurteilen und Beraten, individuelle Weiterbildung und kollegiale Schulentwicklung, Selbststeuerungsfähigkeit im Umgang mit beruflichen Belastungen etc.) über möglichst hohe bzw. entwickelte Kompetenzen und zweckdienliche Haltungen verfügt, die anhand der Bezeichnung „professionelle Handlungskompetenzen“ zusammengefasst werden.

Der kompetenztheoretische Professionalisierungsansatz fokussiert damit die populationsbeschreibende Kompetenzmessung unter Beizug standardisiert-statistischer Methoden mit dem Ziel der Standardisierung von Lehrpersonenbildungssystemen (Bölsterli et al., 2010, S. 277). Wissens- und Kompetenzbereiche werden nicht nur deduktiv, sondern auch auf der Basis empirischer Arbeiten festgelegt, die spezifische berufsbezogene Fähigkeiten von Lehrpersonen mit den Lernerfolgen der Schülerinnen und Schüler in Beziehung setzen (Terhart, 2011). Dieser Ansatz basiert insbesondere auf der Expertiseforschung und der Taxonomie der für den Lehrerberuf relevanten Wissensdimensionen nach Shulman (1986) (Cramer, 2012; Järvinen, 2014; Kamm & Bieri, 2008, cf. Kap. I, 2.5.1). Die Frage der Domänenspezifität der Professionalität, die sich in diesem Zusammenhang insbesondere vor dem Hintergrund der Schulfächer bzw. Lehrgebiete stellt, rückt demzufolge vermehrt ins Zentrum der Forschungsbemühungen. Seifried & Ziegler (2009) erwähnen hierzu zwei empirisch belegte Beispiele, das eine aus der Mathematik (Brunner et al., 2006), das andere zu Unterrichtsplanungen (e.g. Terbrügge, 2001, cf. Kap. I, 2.8.3). Sie erachten die empirische Befundlage aber als zu dürftig und betonen, dass auch untersucht werden sollte, welchen Einfluss darüber hinausgehende, berufsfeldspezifische Faktoren wie u.a. die Adressaten oder der Bildungsgang auf die Entwicklung domänenspezifischer Professionalität haben (Seifried & Ziegler, 2009, S. 90). In neuerer Zeit werden Studien domänenspezifischer Art vorangetrieben, in denen die Rolle der Ausbildung und der Aufbau eines professionell relevanten Wissens untersucht werden (Tenorth, 2013). Sie sind Gegenstand diverser

Forschungsvorhaben. Genauere Befunde hierzu werden im zweiten Teil der Arbeit dargelegt (cf. Kap. II, 1.1).

Beiden hier beschriebenen Ansätzen, dem *struktur-* wie auch dem *kompetenztheoretischen* Ansatz, fehlt die empirische Evidenz in Bezug auf die Bedeutung professioneller Kompetenzen für die Qualität von Unterricht sowie den Lernfortschritt und die Persönlichkeitsentwicklung von Schülerinnen und Schülern (Baumert & Kunter, 2006, 2013). Sie werden im wissenschaftlichen Diskurs zur Lehrerprofessionalität weitgehend kontradiktorisch aufgefasst (e.g. Baumert & Kunter, 2006; Helsper, 2007; Tenorth, 2006; Tillmann, 2011). Dennoch enthalten beide Ansätze hinsichtlich der Lehrpersonenausbildung durchaus wichtige Elemente. Der folgende Abschnitt soll diese kurz beleuchten und damit die Kontradiktion etwas entschärfen.

III. Struktur- und kompetenztheoretisches Verständnis von Professionalität – eine Gegenüberstellung

Ogleich ein Widerspruch zwischen dem *struktur-* und dem *kompetenztheoretischen* Verständnis von Professionalität im wissenschaftlichen Diskurs zur Lehrerprofessionalität nicht von der Hand zu weisen ist, lassen sich gemäss Terhart (2011, S. 209) auch „wichtige, zum Teil sogar sich wechselseitig ergänzende, ja bestätigende Erkenntnisse“ in beiden Ansätzen finden. Den ergänzenden Aspekt beider Ansätze verdeutlicht Cramer (2012, S. 33) damit, dass er dem Prozess des kumulativen Kompetenzerwerbs die zentrale Bedeutung der Professionalität zuschreibt, weil dieser in seinen Augen im Rahmen institutionalisierter Lehrpersonenbildung die Ausbildungsmassnahmen überhaupt erst rechtfertigt. Gleichzeitig erachtet er aber die seit der Akademisierung erfolgte Implementation von Schlüsselqualifikationen wie Handlungsfähigkeit in variablen Situationen oder Reflexionsvermögen auch als wesentliche Aspekte der Lehrpersonenbildung (Cramer, 2012, S. 34). Reinisch (2009) sieht dies in gewisser Weise ähnlich. Er setzt Professionalität und Expertise gleich und misst der Expertenforschung die grössere Bedeutung bei (ebd., S. 37). Im professionssoziologischen Ansatz sieht er aber die Inspiration zu anregenden Hypothesen für die Forschung. Gemäss diesen Ausführungen stufen beide Autoren den kompetenztheoretischen Ansatz als wichtiger für die Lehrpersonenbildung ein. Zierer (2015) hingegen plädiert in einer neueren Arbeit für eine Kombination beider Ansätze, da der kompetenztheoretische Ansatz die Professionalität zu stark auf Wissen und Fähigkeiten reduziert und damit die reflexive Herangehensweise fehlt.

In diesem Zusammenhang stellt sich die Frage, welches Professionalitätsverständnis der naturwissenschaftsdidaktischen Forschung und auch der vorliegenden Studie zugrundeliegt. Fachdidaktische Studien in den Naturwissenschaften orientieren sich traditions-

gemäss an den jeweiligen Domänen Physik, Biologie oder Chemie und beziehen sich überwiegend auf den kompetenztheoretischen Professionalisierungsansatz (Bölsterli et al., 2010, S. 276). Der Fokus liegt dabei sehr häufig auf der Expertise von (angehenden) Lehrpersonen, wobei die Erfassung des Professionswissens mit Hilfe neu entwickelter, domänenspezifischer Messinstrumente Gegenstand vieler neuerer Forschungsvorhaben ist (Krauss, Lindl, Schilcher & Tepner, 2017). Auch die vorliegende Studie, die Unterrichtsplanungen angehender Lehrpersonen in Bezug auf das darin beobachtbare fachdidaktische Wissen zur Planung von Experimentalstunden im Sinne der Erkenntnisgewinnung untersucht, stützt sich auf den kompetenztheoretischen und domänenspezifischen Ansatz. Die Vorbereitung und Inszenierung von Unterricht stehen im Zentrum der Untersuchung. Dabei stellen sich Fragen zu möglichen Hintergründen der beobachtbaren Expertise, der eine grosse Bedeutung für gelingenden Unterricht zugeschrieben wird (Hattie, 2013; Lipowsky, 2006). In der Lehr- und Lernforschung haben sich drei wichtige Forschungsströmungen mit diesen Fragen historisch aufeinanderfolgend befasst. Diese sollen zur Einordnung der naturwissenschaftsdidaktischen Forschung und damit auch der vorliegenden Untersuchung im nächsten Abschnitt kurz umrissen werden.

2.3 Paradigmen der empirischen Professionsforschung – Lehrpersonen als zentrale Akteure schulischer Bildungsprozesse

Zu den Merkmalen, die gute Lehrpersonen als zentrale Akteure schulischer Bildungsprozesse charakterisieren, zählt Weinert (1996, S. 141) bestimmte Persönlichkeitsmerkmale, wirksame Lehrtechniken sowie die professionalisierte Unterrichtsexpertise. Damit nimmt er direkt Bezug auf drei Forschungsströmungen, die sich dieser Frage aus verschiedenen Blickwinkeln angenähert haben und die historisch aufeinanderfolgend durch Akzentverschiebung aus der vorangehenden Perspektive entwickelt worden sind (Krauss, 2011). Es sind dies das «Persönlichkeitsparadigma», das «Prozess-Produkt-Paradigma» und das «Expertenparadigma» (e.g. Bromme, 1997; Bromme & Rheinberg, 2006; Krauss, Neubrand, et al., 2008).

2.3.1 Das Persönlichkeitsparadigma – Persönlichkeitsattribute der Lehrperson

In den 1950er und 1960er Jahren hat sich das «Persönlichkeits-Paradigma» mit grössten-teils nicht veränderbaren und relativ stabilen Personenmerkmalen wie u.a. Charaktereigenschaften und Tugenden befasst (Gräsel & Trempler, 2017, S. 2), um damit Unterschiede des „Lehrerfolges“ (Bromme & Rheinberg, 2006, S. 299) erklären und vorhersagen zu

können. Untersuchungen zur Prüfung dieser Annahme haben jedoch eher geringe, z.T. triviale und inkonsistente Korrelationen zwischen dem Verhalten und den Charaktereigenschaften der Lehrpersonen und ihrer Wirkung auf das Lernen der Schülerinnen und Schüler ergeben (e.g. Bromme, 1997; Gräsel & Trempler, 2017; Helmke, 2012; Krauss & Bruckmaier, 2014, S. 241). Es blieb unklar, ob die definierten Eigenschaften tatsächlich in einen Zusammenhang mit dem Lernerfolg von Lernenden gebracht werden können (Mayr, 2011, S. 126). Daher liess sich mit diesem Ansatz auch kein generalisiertes Persönlichkeitsprofil des „guten“ und „erfolgreichen“ Lehrers aus den Befunden ableiten (Bromme & Rheinberg, 2006, S. 299). Allerdings haben die Ergebnisse darauf hingedeutet, dass erlernbare Kompetenzen und erlernbares Wissen sowie konkrete Situationen (z.B. die Unterrichtssituationen) einen grösseren Einfluss auf die Qualität des Unterrichts haben könnten (Mayr, 2011, S. 125f.). Unter dem Einfluss des Behaviorismus rückte daher etwa ab 1960 die systematische Untersuchung des Unterrichts in den Mittelpunkt (Krauss & Bruckmaier, 2014). Das sogenannte «Prozess-Produkt-Paradigma» wird im nächsten Abschnitt skizziert.

2.3.2 Das Prozess-Produkt-Paradigma – das Lehrerhandeln und seine Wirkung

Mit dem «Prozess-Produkt-Paradigma» wechselte die wissenschaftliche Aufmerksamkeit von der Person mit ihren Persönlichkeitsattributen zum konkreten Lehrerhandeln, d.h. zum Verhalten der Lehrperson im Unterricht. Das Modell postulierte die Wirkung stabiler Merkmale iniziiert Unterrichtsprozesse durch die Lehrperson auf Lerneffekte bei den Schülerinnen und Schülern. Die Wirkung (überwiegend Korrelationen) einzelner Verhaltensmuster und Fertigkeiten (Prozesse) der Lehrenden wurde dabei in Bezug auf eng umschriebene und sichtbare Indikatoren des Schülerverhaltens wie Lernergebnisse (Produkt) analysiert (Bromme & Rheinberg, 2006). Die Produkte umfassten kognitive, aber auch motivational-affektive und metakognitive Aspekte des Lernens. Voraussetzungen der Lernenden (z.B. Vorwissen, Vorerfahrungen der Lernenden) sowie Kontextvariable (elterlicher Einfluss, Einfluss durch Gleichaltrige etc.) wurden als moderierende Variable mitberücksichtigt (Seidel, 2014a, S. 783). Ergebnisse aus diesen Forschungsbemühungen sind keineswegs einheitlich (Besser & Krauss, 2009). Sie haben aber eine Vielzahl empirischer Befunde über lern- und leistungsrelevante Merkmale des Unterrichtens geliefert, welche die Bedeutung der Unterrichtsprozesse unterstrichen (Ditton, 2009; Helmke, 2012). „Aufgrund der einseitigen Untersuchung der Relation zwischen unterrichtlichen Handlungsweisen und den Lernergebnissen (als Output)“ ist die Prozess-Produkt-Forschung jedoch auch kritisiert worden (Seidel, 2014a, S. 784). Zum einen sind die Lernenden nicht als aktive Akteure im Unterricht berücksichtigt und die interaktiven Elemente zwischen ihren

Voraussetzungen und dem Lehrerhandeln nicht miteinbezogen worden. Zum anderen sind die Variablen gegenüber dem Fachinhalt weitgehend unabhängig operationalisiert worden (Bromme & Rheinberg, 2006, S. 302; Shulman, 2015). Auf der Grundlage von Erkenntnissen der Kognitionsforschung sowie konstruktivistischer Sichtweisen des Wissenserwerbes ist das «Prozess-Produkt-Paradigma» zu einem «Prozess-Mediations-Produkt-Paradigma» erweitert worden. Dabei ist die Interaktion zwischen den Merkmalen der Lernenden und den Handlungen der Lehrperson aufgegriffen und in entsprechenden theoretischen Rahmenmodellen berücksichtigt worden. Der Unterricht wird in diesen Modellen als Angebot aufgefasst, das von den Lernenden auf der Basis ihrer Voraussetzungen entsprechend genutzt werden kann. Auch werden die individuellen Informationsverarbeitungsprozesse der Schülerinnen und Schüler (Mediationen) bei der Untersuchung von Lehr- und Lernprozessen miteinbezogen (Besser & Krauss, 2009; Fend, 1998; Helmke, 2012; Seidel, 2014b; Seidel & Reiss, 2014) (cf. Kap. I, 2.6.1). Mit dieser Erweiterung der Mediations- und Interaktionseffekte hat sich die Perspektive von der kollektiven hin zu einer individuellen Sicht verschoben (Brunner et al., 2006). Da das Gelingen von Unterricht aber in erheblichem Masse vom Kontext der Unterrichtsstunde (Fach, Thema oder auch Klassenzusammensetzung) abhängig ist (Besser & Krauss, 2009) und es professionellen Wissens bedarf, um angemessene Aufgaben stellen zu können (Bromme & Rheinberg, 2006), rückte auch die Lehrperson als Experte/-in für die Unterrichtsgestaltung wieder in den Mittelpunkt des Forschungsinteresses. Das sogenannte «Expertenparadigma» wird daher im nächsten Abschnitt näher erläutert.

2.3.3 Das Expertenparadigma – Lehrpersonen als Expertinnen und Experten für das Lernen und Lehren

Beeinflusst durch die kognitive Wende und die damit verbundene Expertiseforschung entstand ab den 1980er Jahren das sogenannte «Expertenparadigma», eine Synthese aus dem «Persönlichkeitsparadigma» und dem «Prozess-Produkt-Paradigma» (Cauet, 2016). Im Zentrum dieses Paradigmas und Forschungsansatzes stehen nicht mehr die stabilen angeborenen Charaktereigenschaften, sondern vielmehr spezifische Fähigkeiten und Fertigkeiten der Lehrpersonen, von denen angenommen wird, dass sie im Rahmen einer speziellen und langjährigen Ausbildung erlernt werden können (cf. Gruehn, 2000, S. 27). In diesem Zusammenhang lassen sich zwei Ansätze von Expertise unterscheiden. Zum einen der leistungsorientierte Ansatz, der die Lehrperson als Problemlöser definiert, zum anderen der wissensorientierte Ansatz, der das Wissen von Personen im Sinne von Expertenwissen für den beruflichen Erfolg in den Fokus rückt (Gräsel & Templer, 2017; Krauss, 2011; Krauss & Bruckmaier, 2014). Letzterer ist durch Bromme (1997) auf den Lehrerberuf übertragen worden. Diese Anwendung hat in der Forschung Resonanz hervorgerufen, weshalb auch

vom «Expertenparadigma» in der Lehrer- und Unterrichtsforschung die Rede ist (König, 2010). Die Lehrpersonen werden in dieser professionssoziologisch orientierten Perspektive als „Experten für das Lernen und Lehren in der Schule betrachtet“ (Bromme, 2008, S. 159). Lehrerexpertise ist somit berufsbezogenes Wissen und Können, das den Lehrerberuf zu einer Profession macht (ebd.) und sich auf professionelles Wissen oder Professionswissen stützt, das es braucht, um die beruflichen Aufgaben zu erfüllen (Baumert et al., 2011). Das Wissen und Können von Experten kann aber immer nur im Zusammenhang mit den zu bewältigenden Anforderungen verstanden werden (Bromme, 2008, S. 162). Eine systemische Sicht auf das Unterrichtsgeschehen mit Bezug zum Lehrerhandeln sowie zu subjektiven Theorien oder Überzeugungssystemen von Lehrpersonen ist daher auch weiterhin erforderlich (Krauss & Bruckmaier, 2014, S. 242). Als Gütemassstab gilt dabei die zuverlässige und dauerhafte Bewältigung der beruflichen Anforderungen (Oser, 2009). Im Fokus dieses Forschungsansatzes stehen vergleichende Analysen von Unterricht und Wissen erfolgreicher bzw. weniger erfolgreicher oder erfahrener bzw. unerfahrener Lehrpersonen. Zudem interessiert die Untersuchung des Wissenserwerbs von Lehrpersonen, insbesondere von angehenden Lehrpersonen während ihrer Ausbildung. Für die Profession relevante Kompetenz- und Wissensdimensionen werden dazu theoretisch, aber auch empirisch hergeleitet und untersucht, ohne dabei die situative Unsicherheit des Lehrerhandelns, die keine vollständige Standardisierbarkeit und Technisierung zulässt, ausser Acht zu lassen (Baumert & Kunter, 2006).

Das «Expertenparadigma» wird derzeit neben dem «Prozess-Mediations-Produkt-Paradigma» in der pädagogischen Psychologie favorisiert und nimmt in der empirischen Forschung zum Lehrerberuf eine wichtige Stellung ein. Ergänzend dazu sind auch Persönlichkeitsmerkmale von Lehrpersonen Gegenstand neuerer Untersuchungen. Diese werden aber eher als Bedingungen angesehen, die den Verlauf des Studiums von angehenden Lehrpersonen oder den beruflichen Alltag von Lehrenden begünstigen oder beeinträchtigen können (Gebauer, 2013).

Aufgrund der Einführung der Bildungsstandards und der damit verbundenen Output- bzw. Kompetenzorientierung sind auch in der naturwissenschaftsdidaktischen Forschung die Lehrpersonen und deren Expertise ins Zentrum des Forschungsinteresses gerückt. Die empirische Professionsforschung der Naturwissenschaftsdidaktik stützt sich daher aktuell mehrheitlich auf das «Expertenparadigma» (Ergönenc, Neumann & Fischer, 2014; Gess-Newsome et al., 2017; Grossschedl, Harms, Kleickmann & Glowinski, 2015; Jüttner & Neuhaus, 2013; Stender, Brückmann & Neumann, 2017). Dabei wird dem Unterricht und seiner Gestaltung für die Entwicklung und für den systematischen Erwerb naturwissenschaftlicher Kompetenzen eine wesentliche Bedeutung beigemessen (Prenzel & Seidel,

2008). Es ist allerdings nach wie vor nicht geklärt, welche professionelle Handlungskompetenz und welches Professionswissen Lehrpersonen für die Planung und Durchführung qualitätsvollen naturwissenschaftlichen Unterricht benötigen, insbesondere für den Unterricht zum Experimentieren im Dienst naturwissenschaftlichen Erkenntnisgewinns im Fach Biologie.

Das nächste Kapitel befasst sich vor diesem Hintergrund mit der professionellen Handlungskompetenz von Lehrpersonen. In einem ersten Schritt erfolgt eine begriffliche Annäherung an den Kompetenzbegriff in seiner ganzen Unschärfe (Kap. I, 2.4.1). Auch werden verschiedene Herangehensweisen bei der Modellierung von Kompetenzen und damit verbunden auch verschiedene Modelltypen vorgestellt (Kap. I, 2.4.2). Schliesslich werden ausgewählte Modellierungsansätze professioneller Handlungskompetenz beschrieben und einander gegenübergestellt (Kap. I, 2.4.3). Dabei wird insbesondere auf die für die vorliegende Untersuchung zentrale Handlungsanforderung der Unterrichtsplanung Bezug genommen und soweit möglich, auch das Experimentieren im Sinne der Erkenntnisgewinnung berücksichtigt.

2.4 Professionelle Handlungskompetenz von Lehrpersonen – Modellierungsansätze der Professionsforschung

Aktuelle Studien geben Anlass zur Annahme, dass Professionalität im Handeln und damit die Handlungskompetenz von Lehrpersonen im Sinne einer Wirkkette aufgefasst werden können. So beeinflussen die Qualität der im Verlauf der Lehrerinnen- und Lehrerbildung erworbenen professionellen Fähigkeiten und sich daraus entwickelnden prozeduralen Routinen das unterrichtliche Handeln von Lehrpersonen (e.g. Baumert & Kunter, 2011a; Hattie & Yates, 2015; Künsting, Billich & Lipowsky, 2009; Lipowsky et al., 2009; Pauli & Reusser, 2009; Voss et al., 2011; Wagner et al., 2016; Zlatkin-Troitschanskaia, Beck, Sembill, Nickolaus & Mulder, 2009a). Dies wiederum zeigt Wirkung auf die Leistungen der Schülerinnen und Schüler (Kunter et al., 2013; Lipowsky et al., 2009; Weinert, 1990) sowie auch auf deren Motivation und Interesse (Krapp & Prenzel, 2011). Dieses Professionalisierungsmodell im Sinne einer Kausalkette ist zwar nicht unbestritten (cf. Cochran-Smith, 2001), erlaubt aber eine Groborientierung und knüpft an Grundelemente der Lehrpersonenbildung an (Frey, 2014, S. 712). Hinsichtlich der Modellierung von Handlungskompetenz sowie der zentralen Merkmale, mit welchen kompetente Lehrpersonen beschrieben werden sollen, lässt sich in der Fachliteratur aber nach wie vor kein Konsens ausmachen (Baumert & Kunter, 2006). Ein theoretisches Rahmenmodell fehlt (ebd.). Dass es sich bei den untersuchten Merkmalen um berufsspezifische, veränderbare Dispositionen handelt, die Lehrpersonen zur erfolgreichen Ausübung ihres Berufes benötigen, scheint aber Grundidee aller

Ansätze zu sein (Baumert & Kunter, 2006). Die folgenden Abschnitte befassen sich vor diesem Hintergrund mit den berufsspezifischen Dispositionen und dem Kompetenzbegriff, aber auch mit Kompetenzmodellierungen und Modellen professioneller Handlungskompetenz.

2.4.1 Berufsspezifische Dispositionen – eine Annäherung an den Kompetenzbegriff

Berufsspezifische Dispositionen werden je nach Forschungshintergrund unterschiedlich definiert. In der Berufsbildung werden diese Dispositionen als intellektuelle Fähigkeiten, Vorwissen, Fertigkeiten und Routinen, motivationale Merkmale, metakognitive und volitionale Kontrollsysteme sowie persönliche Wertorientierungen in einem komplexen handlungsregulierenden System aufgefasst (Seeber & Nickolaus, 2010, S. 11). Sie gelten als Voraussetzungen, um „anstehende Probleme zielorientiert und verantwortungsvoll zu lösen, zu reflektieren und zu bewerten und das ganze eigene Repertoire an Handlungsmustern weiterzuentwickeln“ (Frey, 2006, S. 31). Es handelt sich hierbei um einen holistischen Ansatz zur Erfassung einer multikriterialen beruflichen Handlungskompetenz. Dieser liefert aber bisher keine verlässlichen Ergebnisse (Seeber & Nickolaus, 2010). Projekte der Lehrpersonenbildungsforschung stützen sich demgegenüber meist auf ein engeres Verständnis beruflicher Dispositionen. Sie berufen sich mehrheitlich auf den Kompetenzbegriff von Weinert (2001a), in Abgrenzung zu generalisierten, kontextunabhängigen Leistungskonzepten (wie z.B. Intelligenz, überfachliche Kompetenzen im Sinne von Schlüsselqualifikationen). Weinert (2001a, S. 27f.) versteht unter Kompetenzen

[...] die bei Individuen verfügbaren oder durch sie erlernbaren kognitiven Fähigkeiten und Fertigkeiten, um bestimmte Probleme zu lösen, sowie die damit verbundenen motivationalen, volitionalen und sozialen Bereitschaften und Fähigkeiten, um die Problemlösungen in variablen Situationen erfolgreich und verantwortungsvoll zu nutzen.

Nach diesem Verständnis beschreibt Kompetenz die erlernbaren oder vermittelbaren persönlichen Voraussetzungen zur erfolgreichen Bewältigung situationsspezifischer Anforderungen und Aufgaben. Dabei werden kognitive und affektive Dispositionen von Lehrpersonen als Voraussetzungen für kompetentes Handeln im Unterricht erachtet (Baumert & Kunter, 2011b; Hartig & Klieme, 2006; Ohle, Fischer & Kauertz, 2011). Die Kompetenz wird demzufolge als Disposition interpretiert, die sich aber erst in der situativen Bewältigung von Anforderungen, d.h. in der „Performanz“ des Handelns zeigt (Klieme & Hartig, 2007, S. 13). Das bedeutet, dass sich Kompetenz in gewisser Weise nur über eine „Menge hinreichend ähnlicher realer Situationen, in denen bestimmte, ähnliche Anforderungen

bewältigt werden müssen“, definieren lässt (Hartig, 2008, S. 21). Die Kompetenz ist demnach immer kontextabhängig und damit auch „more closely related to real life“ (Koeppen, Hartig, Klieme & Leutner, 2008). Im Rahmen der naturwissenschaftsdidaktischen Expertiseforschung wird sie deshalb auch häufig domänenspezifisch beschrieben und operationalisiert (Klieme & Leutner, 2006; cf. Koeppen et al., 2008; Weinert, 2001b).

Der Begriff der Kompetenz bezieht sich so gesehen sowohl auf die Handlungsvollzüge als auch auf die ihnen zugrundeliegenden mentalen Prozesse und Kapazitäten (Hartig & Klieme, 2006, S. 13). Diese Dichotomie des Konstruktes versuchen Blömeke, Gustafsson und Shavelson (2015) mit einer erweiterten Definition des Begriffs aufzunehmen, die lautet: “Competencies are the latent cognitive and affective-motivational underpinning of domain-specific performance in varying situations” (Blömeke et al., 2015). Sie unterscheiden dabei zwischen *competencies* und *competences* und skizzieren damit zwei verschiedene Herangehensweisen zur Erfassung von Kompetenz (Blömeke et al., 2015, S. 6). Unter *competencies* subsumieren sie die wesentlichen kognitiven und affektiven Dispositionen, die zur Lösung von Problemen nötig sind. Es geht hier also um eine eher analytische Haltung mit dem Ziel, Dispositionen zu erfassen, die einem bestimmten Verhalten zugrunde liegen und mit diesem in Verbindung gebracht werden können. Demgegenüber umschreiben sie mit dem Begriff *competences* ein eher holistischeres Vorgehen, mit dem über Verhaltensbeobachtungen in Realsituationen die Performanz erfasst wird und Rückschlüsse auf die Kompetenz bzw. Kompetenzfacetten gezogen werden. Beide Herangehensweisen lassen ausser Acht, dass sich Dispositionen wie auch die Performanz situationsspezifisch verändern können. Maag Merki (2004, S. 204) weist darauf hin, dass der Zusammenhang zwischen Kompetenz im Sinne von Handlungsvoraussetzungen und Performanz, d.h. der Kompetenz, die sich in der Handlung zeigt, nicht deterministisch aufgefasst werden kann, da das Handeln immer auch durch Merkmale der konkreten Situation und durch die individuumsspezifische Deutung beeinflusst wird. Oser (2013) teilt diese Beobachtung und erwähnt in diesem Zusammenhang eine zusätzliche Prozessdimension, durch die sich die Kompetenz in der Situation erst konstituiert. Diese Dichotomie der Herangehensweisen sowie die Prozessdimension werden im neuen Modell von Blömeke et al. (2015) aufgegriffen (Abb. 3). Kompetenz wird hier als Prozess im Sinne eines horizontalen Kontinuums mit vielen einzelnen Schritten skizziert, mit der Empfehlung, unabhängig von der jeweiligen Herangehensweise zur Erfassung der Kompetenz, den Prozess mit in Betracht zu ziehen. Dieser verbindet die kognitiven und affektiven Ressourcen (Dispositionen) einer Lehrperson mit dem beobachtbaren Verhalten.

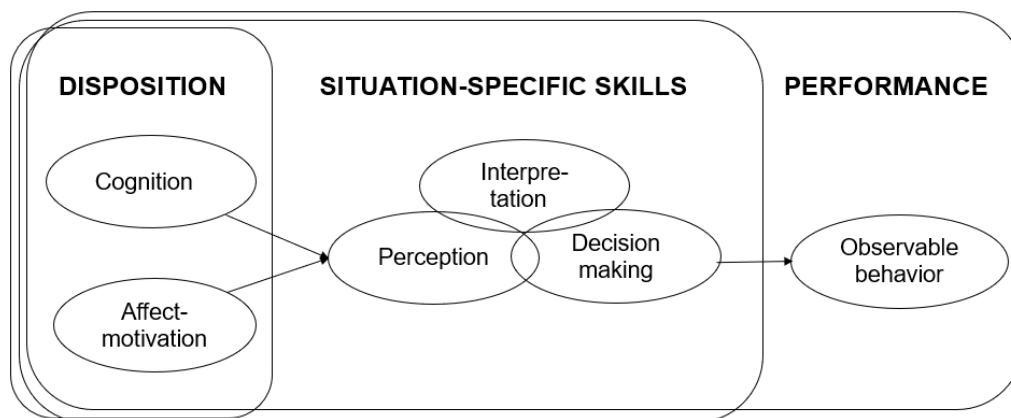


Abbildung 3: Kompetenzmodellierung als Kontinuum (Blömeke et al., 2015)

Als Mediatoren zwischen beiden Polen werden situationsspezifische und verhaltensnahe *skills* wie Wahrnehmung von lernrelevanten Situationen (*perception*), deren Analyse und Interpretation (*interpretation*) und die anschließende Entscheidung über Handlungsoptionen (*decision making*) angenommen (Blömeke et al., 2015, S. 8).

Das «Prozessmodell» zum Kompetenzbegriff von Blömeke et al. (2015) verdeutlicht aufgrund seiner vielfältigen Facetten menschlichen Denkens, Handelns und Empfindens die Komplexität dieses Konstruktes, das bislang einer „scharfen Operationalisierung widerstrebt“ (Gut-Glanzmann, 2012, S. 13). Gemäss Gut-Glanzmann sind die Gründe dafür in den Abhängigkeiten zwischen Kompetenzbegriff und konkreten Aufgaben zu suchen (2012, S. 21ff.):

1. Die Beschreibung einer zu bewältigenden Aufgabe ist stets integraler Bestandteil jeder Kompetenzdefinition.
2. Aufgabenbeispiele und -merkmale sowie entsprechende Kodierschemen sind Grundlage für die Explikation von Kompetenzbeschreibungen.
3. Authentische Kompetenzausübungen in Realsituationen unterscheiden sich von erzwungenen experimentellen Kompetenzmessungen in der Kompetenzdefinition und Aufgabe.

Diese Gründe führen dazu, dass in der Praxis auf Modellierungen von Aufgaben, die in Realsituationen ausgeübt werden sollen, oft verzichtet wird. Dafür werden experimentelle Aufgaben verfolgt, die mit geringerem Aufwand durchgeführt werden können und methodisch erfolgsversprechender sind (Gut-Glanzmann, 2012). Dies hat auch Anpassungen in Bezug auf Kompetenzbeschreibungen bzw. -definitionen zur Folge. Weinert empfiehlt aufgrund theoretischer und pragmatischer Überlegungen die kognitiven Dispositionen getrennt von den affektiven Dispositionen zu erfassen, um Wechselwirkungen analytisch darstellen zu können (cf. Klieme & Hartig, 2007; Koeppen et al., 2008; Weinert,

2001a). Klieme und Leutner (2006) greifen diese Eingrenzung des Kompetenzbegriffs auf und beschreiben Kompetenzen als „kontextspezifische kognitive Leistungsdispositionen, die sich funktional auf Situationen und Anforderungen in bestimmten Domänen beziehen“ (Klieme, & Leutner, 2006, S. 879). Dieser auf die kognitiven Dispositionen eingeschränkte Kompetenzbegriff findet Anwendung im Rahmen des DFG²-Schwerpunktprogramms «Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen».

In der Literatur naturwissenschaftsdidaktischer Forschung finden sich zu beiden vorangehend beschriebenen Herangehensweisen Studien. Studien, die auf dem Kompetenzbegriff von Weinert (2001a) und Dispositionen basieren, von denen angenommen wird, dass sie bedeutenden Einfluss auf das Lehrerhandeln ausüben, haben jedoch schon länger Tradition. Kern vieler dieser Studien ist die Untersuchung von kognitiven Dispositionen (Aufschnaiter & Rogge, 2010; Hasse, Joachim, Bögeholz & Hammann, 2014; Ohle et al., 2011; Riese, 2009). Weinerts (2001b) Forderung, auch handlungsorientierte Komponenten wie die motivationalen und volitionalen Dispositionen bei der Erfassung von Kompetenzen miteinzubeziehen, wird jedoch trotz der vorherrschenden Anerkennung deren Bedeutung wenig berücksichtigt (Schecker & Parchmann, 2006, S. 46). Als mögliche Gründe lassen sich dafür Operationalisierungsfragen zur genauen Erfassung affektiver Kompetenzen, aber auch testökonomische Überlegungen annehmen. Letzteres ist auch der Grund, weshalb Untersuchungen zur Performanz, d.h. zum Lehrerhandeln erst mit den neueren technologischen Möglichkeiten in der naturwissenschaftsdidaktischen Forschung zugenommen haben (Riegel, 2013). Es finden sich hierzu Untersuchungen zu allen drei naturwissenschaftlichen Fächern (cf. Brückmann & Bernholt, 2013; Seidel et al., 2006; Wadouh, Sandmann & Neuhaus, 2009). Untersuchungen von Planungen sind hier aber vergleichsweise selten (König et al., 2015; Seel, 1997, 2011).

Unabhängig von der Herangehensweise für die empirische Erfassung domänenspezifisch beschriebener Kompetenz zur Bewältigung komplexer Handlungsanforderungen von Lehrpersonen, stellt sich die Frage nach der Modellierung, mit der die Anforderungen festgelegt und auf deren Grundlage entsprechende Messinstrumente entwickelt werden sollen (cf. Hartig & Klieme, 2006). Der nächstfolgende Abschnitt beleuchtet häufig diskutierte Vorgehensweisen der Modellierung, verschiedene Modelltypen und Methoden der Kompetenzmessung.

² Deutsche Forschungsgemeinschaft

2.4.2 Kompetenzmodellierungen in der Lehrpersonenbildungsforschung – Vorgehensweisen, Modelltypen und Messmethoden

Die Modellierung und empirische Erfassung anforderungs- bzw. domänenspezifisch beschriebener Kompetenzen oder Kompetenzfacetten erfordert eine Systematik in der Vorgehensweise, insbesondere weil Kompetenzen nicht direkt beobachtbar sind, sondern sich nur über die Performanz erschliessen lassen (Shavelson, 2013). Es bedarf dafür tragfähiger und für die Messung und das Lernen umsetzbarer Kompetenzmodelle (Schecker & Parchmann, 2006, S. 47). Je nach der verfolgten Zielsetzung der Modellierung, den theoretischen Grundlagen, auf denen das Modell basiert sowie den gewählten methodischen Modellierungsvarianten können im Rahmen einer Domäne aber unterschiedliche Modelle resultieren (Schaper, Hilligus & Reinhold, 2009, S. 2). Die verschiedenen Ansätze zur Modellierung von Kompetenzen sowie die daraus abgeleiteten verschiedenen Typen von Kompetenzmodellen sollen in der hier erwähnten Reihenfolge kurz umrissen werden.

I. Kompetenzmodellierungen

Kompetenzmodellierungen können sich aufgrund theoretischer Grundlagen, aber auch weiterer Strukturierungselemente unterscheiden. Dazu gehören der gewählte Fokus in Bezug auf den Tätigkeitsausschnitt, das Spektrum der betrachteten Kompetenzfacetten, die Vorgehensweise bei der Modellierung, aber auch die Fragestellungen, die verfolgt werden (Schaper, 2009, S. 170). Der Tätigkeitsausschnitt und damit der inhaltliche Fokus der modellierten Kompetenz mit der damit verbundenen Aufgabe kann sehr breit und umfassend auf ein gesamtes Tätigkeitsfeld einer Domäne, oder aber nur auf ausgewählte Teilbereiche der Lehrertätigkeit ausgerichtet sein. Im ersten Fall wird ein entsprechend breites Spektrum von Kompetenzfacetten einbezogen, während im zweiten Fall der Fokus nur auf Teilfacetten des Kompetenzkonstrukts liegt (ebd.). Die Vorgehensweise bei der Kompetenzmodellierung ist ein weiteres Strukturierungselement. Sie kann *induktiv* oder *deduktiv* mit jeweils unterschiedlichen Ansätzen sein (Schaper et al., 2009). Ansätze der induktiven Vorgehensweise fassen hierbei auf einer empirischen Analyse der Tätigkeiten in einem Fachgebiet und definieren Kompetenzfacetten unter Berücksichtigung realer Situationen und Anforderungen (Frey, 2014, S. 713). Dies geschieht beispielsweise mit Hilfe einer standardisierten Expertenbefragung. Demgegenüber gehen Ansätze der deduktiven Vorgehensweise der Modellierung von bildungstheoretisch oder pragmatisch begründeten Konzepten zu relevanten Bildungszielen aus und versuchen bestehende Kompetenzkategorien des aktuellen Forschungsstands empirisch abzubilden (cf. Frey, 2014; Schaper, 2009). Dies kann sowohl deskriptiv im Sinne einer Interpretation, Rekonstruktion oder Beschreibung, als auch

normativ, im Hinblick auf die erforderlichen Handlungsressourcen, erfolgen (Schecker & Parchmann, 2006). Ein weiteres Strukturierungselement ist auch die Ausrichtung der Fragestellungen, sei es, dass diese eher kompetenzmodellierend oder zusammenhangsanalytischer Natur sind oder aber, dass sie mess- und evaluationsorientierte Zielsetzungen verfolgen. Unabhängig von Zielen und Vorgehensweisen, stellt sich jedenfalls die Frage, welche Teilkompetenzen definiert und in welcher Ausprägung diese beschrieben werden sollen. Die Feingliedrigkeit (Teilkompetenzen) und die Feinkörnigkeit (Ausprägung) bestimmen dabei stets den Auflösungsgrad des Modells (Neumann, 2013, S. 35).

II. Kompetenzmodelle

Was die Modelltypen betrifft, so wird in der Literatur zwischen sogenannten *Kompetenzstrukturmodellen*, *-niveaumodellen* und *-entwicklungsmodellen* unterschieden (cf. Hartig & Klieme, 2006; Schaper, 2009; Schecker & Parchmann, 2006). Bei diesen drei Modelltypen stehen jeweils unterschiedliche Beschreibungsaspekte der Kompetenzkonstrukte und Funktionen der Modelle im Vordergrund (Schaper, 2009, S. 173), wobei Gut in diesem Zusammenhang anmerkt, dass Kompetenzmodelle gleichzeitig auch mehrere Funktionen erfüllen können (Gut-Glanzmann, 2012, S. 19). Diese drei Modelltypen sollen im Folgenden mit Blick auf die vorliegende Studie näher erläutert werden.

Kompetenzstrukturmodelle beschäftigen sich mit den Dimensionen, die den Kompetenzen zugrunde liegen und in diesem Zusammenhang mit der Frage, wieviele und welche Kompetenzdimensionen in einem spezifischen Bereich differenzierbar sind (Hartig & Klieme, 2006; Mayer & Wellnitz, 2014). Auch legen sie die Kompetenzen oder Kompetenzfacetten und die damit korrespondierenden Fähigkeiten in ihrer Ausprägung qualitativ fest. Die Ausprägung kann durch eine Ordinalskala repräsentiert sein, mit der Niveau und Wertigkeit durch die Qualität der darin erreichten Expertise festgelegt werden (Schecker & Parchmann, 2006). Im Zentrum von Kompetenzstrukturmodellen steht eine möglichst umfassende Beschreibung erforderlicher Kompetenzen für eine Domäne und die Gewährleistung einer differenzierten Kompetenz- und Leistungsmessung (Schaper, 2009, S. 174).

Kompetenzniveaumodelle beschreiben im Vergleich dazu die spezifischen Fähigkeiten bei unterschiedlichen Ausprägungen einzelner Kompetenzen. Es handelt sich hierbei um Kompetenzcharakterisierungen, die „auf der Analyse der schwierigkeitsbestimmenden Merkmale der jeweils gelösten Aufgaben in einem Kompetenztest beruhen“ (Schaper, 2009, S. 275). Dabei werden meist die Itemschwierigkeiten und Personenfähigkeiten raschskaliert (Schecker & Parchmann, 2006, S. 51). Ziel solcher Modelle ist eine kriteriumsorientierte Beurteilung und Klassifizierung von Testwerten zur Einordnung von Bildungsprozessen (cf. Schaper, 2009).

Kompetenzentwicklungsmodelle bilden den Ablauf von Kompetenzerwerbsprozessen ab und beschreiben auf der Basis angenommener Entwicklungsstufen die Reihenfolge, in welcher eine Person die Kompetenz erwirbt. Allerdings muss bei der Konzeptionalisierung dieses Modelltyps empirisch geprüft werden, ob die gewählten Niveaus auch von ihrer Entwicklung und Aneignung her logisch interpretiert werden können. Diese Modelle dienen vor allem dazu, die Entwicklungszustände beim Kompetenzerwerb zu bewerten und einzuordnen und gegebenenfalls mögliche Lern- und Entwicklungsschritte davon abzuleiten (Schaper, 2009).

Von den hier beschriebenen drei Modelltypen sind in der Lehrpersonenbildungsforschung vor allem Kompetenzstrukturmodelle besonders häufig (Schaper, 2009). Diese sind bisher normativ, empirisch, aber auch mit Hilfe kombinierter Strategien entwickelt worden. Sie besitzen, unabhängig davon, auf welcher Grundlage sie konstruiert wurden, stets hypothetischen Charakter und bedürfen daher empirischer Schritte für ihre Überprüfung und Validierung (Schaper et al., 2009, S. 3). Inwieweit die Modellierung und Erfassung von Kompetenzen von Lehrpersonen allerdings valide gelingen kann und wie Kompetenzen möglichst unterrichts- und handlungsrelevant erfasst werden können (Riese & Reinhold, 2012a), hängt auch von der Wahl der adäquaten Methode der Kompetenzmessung ab. Mögliche Methoden zur Erfassung und Messung der Kompetenzen werden deshalb im nächstfolgenden Abschnitt, im Sinne eines kurzen Überblicks, besprochen.

III. Methoden der Kompetenzmessung

Die Kompetenzforschung in der Lehrpersonenbildung greift methodisch auf ein breites Spektrum an Erhebungsinstrumenten zurück (Riegel, 2013). Dabei variieren die Erhebungsinstrumente in Abhängigkeit von der jeweiligen Erhebungssituation (Grundlagenforschung, Evaluation, individuelle Diagnose), dem jeweils zugrundegelegten Kompetenzverständnis und dem Ziel der Erhebung (Kaufhold, 2006, S. 126). Meist kommen dabei standardisierte Verfahren zur Anwendung (Koeppen et al., 2008). Für fachlich-methodisch orientierte Kompetenzen werden häufig objektive, quantitative Verfahren auf der Basis von Fremdeinschätzungen durch eine oder mehrere Personen gewählt (Sauter & Staudt, 2016, S. 10). Diese reichen von schriftlichen und mündlichen Befragungen, über Tests bis hin zu Beobachtungen und sogenannten Arbeitsproben (Kaufhold, 2006). Letztere entsprechen einer Simulation wichtiger, standardisierter Ausschnitte der Arbeitstätigkeit. Dabei wird das Verhalten der Probanden in diesen Situationen sowie das Arbeitsergebnis einer systematischen Beobachtung unterzogen (Kanning, 2004, S. 425). Mit Hilfe von Arbeitsproben lassen sich somit Fähigkeiten und Fertigkeiten sowie Wissen in Bezug auf den Handlungsprozess erfassen (Kaufhold, 2006). Die Methode der Arbeitsprobe wie auch die Methode

der Beobachtung fokussieren daher handlungsrelevante Kompetenzen oder „*competences*“ und zeichnen sich durch ihre Performanznähe aus. Entsprechend ist die ökologische Validität dieser Verfahren hoch. Demgegenüber zielen die Methoden der Befragung und auch der Tests auf die Erhebung von „*competencies*“ und damit auf die Erfassung von Dispositionen, die einem bestimmten Verhalten zugrunde liegen und mit diesem in Verbindung gebracht werden können (cf. Blömeke, Gustafsson & Shavelson, 2015, Kap. I, 2.4.1). Die ökologische Validität ist hier vergleichsweise geringer. Viele dieser Verfahren werden dank technologischer Neuerungen mittlerweile auch computer- und onlinegestützt durchgeführt. Dieser technische Fortschritt hat auch die videobasierte Unterrichtsforschung, die schon in den 1970er Jahren genutzt worden ist, ab Mitte der 1990er Jahre vorangetrieben (Pauli & Reusser, 2006). Spätestens mit den TIMSS-Videostudien zum Mathematikunterricht (Stigler, Gonzalez, Kawanaka, Knoll & Serrano, 1999) und zum naturwissenschaftlichen Unterricht (Roth et al., 2006), die neue Massstäbe für Videostudien im internationalen Raum gesetzt haben (Brückmann & Duit, 2014), sind Videoanalysen wieder in den Blickpunkt der Lehr-Lernforschung gerückt (Seidel & Prenzel, 2003). Unterrichtsvideos bilden einen interessanten Kontrast zu den bisherigen Erhebungsmethoden und eignen sich in der Kompetenzmessung insbesondere wegen der Ganzheitlichkeit und Authentizität des Datenmaterials, aber auch, weil sie verschiedene Perspektiven und Verfahren der Analyse zulassen (Pauli & Reusser, 2006). Die Videografie in der Lehrpersonenbildungsforschung unterscheidet ebenfalls verschiedene methodische Ansätze. Dazu gehören die sogenannten *video surveys*, die videobasierte Unterrichtsforschung, aber auch Verfahren wie *video-stimulated recalls* und videografierte Unterrichtsbesprechungen (Derry et al., 2010; Pauli & Reusser, 2006; Riegel, 2013). Mit den *video surveys* werden Prozessmerkmale des Unterrichts anhand repräsentativer Stichproben systematisch erfasst und quantifiziert, wie erfolgt bei der TIMSS-Videostudie (Stigler, Gallimore & Hiebert, 2000). Dabei beziehen sich *video surveys* in der Regel auf Sicht- bzw. Oberflächenstrukturen von Unterricht (Kap. I, 2.8.2). Die videobasierte Unterrichtsforschung hingegen ist ein theoriegeleiteter Zugang, bei dem die Analyse videografierten Daten in ein theoretisches Konzept schulischen Lernens eingebettet ist und aus dem beobachtbaren Verhalten Rückschlüsse auf das Konzept gezogen werden. Hier einzuordnen wäre u.a. die IPN-Videostudie zum Physikunterricht (Seidel et al., 2006). Diese beiden methodischen Ansätze zeichnen sich durch ihre Performanznähe und erhöhte ökologische Validität aus und ermöglichen die Erfassung von *competences* (Blömeke et al., 2015). Narrative Verfahren wie *video-stimulated recalls* und videografierte Unterrichtsbesprechungen fokussieren demgegenüber die Analyse von Kompetenzen im Sinn von Dispositionen oder *competencies* und erfassen die subjektiven Theorien der Probandinnen und Probanden. *Stimulated recalls* verwenden dazu einen sogenannten Stimulus (oft in Form von Videovignetten), mit dem Interviewpartnerinnen und

-partner dazu aufgefordert werden, ihre Gedanken explizit zu äussern (Messmer, 2014). Mit diesem Verfahren werden Daten zu präaktionalen Denkprozessen gewonnen (ebd.). Videografierte Unterrichtsbesprechungen ermöglichen im Vergleich dazu Einblicke in subjektiv geprägte, aber unterrichtsrelevante Lernereignisse, die in einer Besprechungssituation geäußert werden (Kreis & Staub, 2012). In beiden narrativen Verfahren, insbesondere bei den *stimulated recalls*, ist die ökologische Validität geringer.

Es zeigt sich, dass die hier beschriebenen methodischen Zugänge zur Erfassung und Messung von Kompetenzen sich ungeachtet ihrer Vorgehensweisen in ihrer Performanznähe und entsprechend auch in ihrer ökologischen Validität unterscheiden (cf. Vogelsang, 2014). Die Ursache dafür liegt u.a. in dem jeweils zugrunde gelegten Kompetenzverständnis und entsprechend in der Kompetenzmodellierung. Vor diesem Hintergrund soll im nächsten Kapitel die Modellierung professioneller Handlungskompetenz näher erörtert werden. Dabei werden verschiedene Modellierungsansätze vorgestellt und die vorliegende Studie darin positioniert.

2.4.3 Professionelle Handlungskompetenz – Modellierungen mit Blick auf den Kompetenzbereich der Unterrichtsplanung

Für die Modellierung professioneller Handlungskompetenz bildeten im deutschsprachigen Raum der Abschlussbericht der Lehrerbildungskommission der Kultusministerkonferenz „Perspektiven der Lehrerbildung in Deutschland“ (Terhart, 2000) und der schweizerische Forschungsbericht „Die Wirksamkeit der Lehrerbildungssysteme“ (Oelkers & Oser, 2000) wichtige Grundlagen. Dabei sind für die Lehrpersonenbildung Modelle mit unterschiedlichen Modellierungsansätzen entwickelt worden. Diese lassen sich in zwei Gruppen einteilen, welche die beiden in Kapitel I, 2.4.1 beschriebenen Herangehensweisen der Kompetenzerfassung widerspiegeln. Das sind auf der einen Seite *standardorientierte Modelle*, die durch die Analyse prototypischer Anforderungssituationen in unterschiedlicher Weise spezifische Dispositionen zum erfolgreichen Handeln im Beruf festlegen (Oser, 2001; Terhart, 2002) und sich durch ihre Performanznähe auszeichnen. Auf der anderen Seite sind es *professionswissensorientierte Modelle*, die normativ spezifische Dispositionen definieren, von denen angenommen wird, dass sie im Zusammenhang mit erfolgreichem Handeln im Unterricht stehen (Baumert & Kunter, 2011a; Bromme, 1997; NBPTS, 2002; Weinert, 2000; Weinert, Schrader & Helmke, 1990).

Ausgewählte Modelle beider Gruppen sollen mit Blick auf die Handlungsanforderung «Planung von Unterricht» und auch auf die Länder, in denen die Untersuchung der vorliegenden Studie stattgefunden hat (Deutschland und Schweiz), genauer betrachtet werden.

I. Standardorientierte Modelle

Standardorientierte Kompetenzmodelle sind auf die Performanz ausgerichtet (Vogelsang, 2014) und legen die sichtbare Problemlösefähigkeit in Arbeits- und Lebenssituationen fest (Wilhelm, 2012). Modellierungsansätze für die Lehrpersonenbildung, in der Schweiz von Oser (2001) und in Deutschland von Terhart (2002), können dieser Gruppe zugeordnet werden. Sie werden in der Folge kurz vorgestellt.

Modell von Oser (2001)

Zum Zweck der Wirksamkeitsevaluation der Lehrpersonenbildung in der Schweiz wurde in einer induktiven Herangehensweise durch Expertenbefragung ein Modell mit Zielvorgaben für die Ausbildung von Lehrpersonen entwickelt (Oser, 2001). Kern dieses Modells von Oser (2001) bilden pädagogisch-psychologisch begründete Standards mit darauf bezogenen Gütemassstäben (Baumert & Kunter, 2006). Die so herausgearbeiteten 88 Standards wurden in 12 Bereiche gruppiert, die eine grosse Bandbreite der Anforderungen abdecken (cf. Frey, 2014):

- Lehrer-Schüler-Beziehung
- Schüler unterstützende Beobachtung und Diagnose
- Bewältigung von Disziplinproblemen und Schülerrisiken
- Aufbau und Förderung von sozialem Verhalten
- Lernstrategien vermitteln und Lernprozesse begleiten
- Gestaltung und Methoden des Unterrichts
- Leistungsmessung
- Medien des Unterrichts
- Zusammenarbeit in der Schule
- Schule und Öffentlichkeit
- Selbstorganisationskompetenz der Lehrperson
- Allgemeine und fachdidaktische Standards

Die 12 Standardgruppen verweisen dabei auf unterschiedliche Referenzpunkte schulischer Handlungsfelder, wobei sich jedoch keine dieser zwölf Klassen mit dem Fachwissen der Lehrperson auseinandersetzt, was auch als möglicherweise gefährliche Fokussierung angesehen werden kann (Wilhelm, 2007, S. 75). Oser (1997) betont dazu, dass es sich bei seinen Standards um Lehrerkompetenzen auf der Handlungsebene handelt, wobei das «Wissen» nur einen notwendigen, aber nicht hinreichenden Bestandteil darstellt. Er erwähnt auch, dass die von ihm entwickelte Standardsammlung keinerlei Anspruch auf Vollständigkeit erhebt. Diesen erachtet er für Standards als unsinnig und wertlos (Oser, 1997,

S. 36). Eine der von Bromme (1997) geäußerten Kernaufgaben, nämlich die Vorbereitung und Planung von Lehr- und Lernprozessen im Unterricht, ist in diesem Modell nicht direkt aufgeführt, aber Teilaspekte davon finden sich in den Standardgruppen, wie u.a.: Schülerunterstützende Beobachtung und Diagnose, Lernstrategien vermitteln und Lernprozesse begleiten, Leistungsmessung sowie allgemeine und fachdidaktische Standards (Oser, 1997, S. 32 ff.). Gemäss Baumert und Kunter (2006) handelt es sich bei diesem Modell um ein generisches, pädagogisch-psychologisches Kompetenzmodell, das anschlussfähig ist an pädagogische und psychologische Theorien, die sich empirisch bewährt haben. Die Spezifität des Fachunterrichts wird jedoch in diesem Modell überhaupt nicht berücksichtigt. Fachdidaktische Kompetenzprofile sind nur als allgemeindidaktische Qualifikationen formuliert und bedürfen daher einer Übersetzung für das jeweilige Fach (Baumert & Kunter, 2006, S. 478). Das Modell von Terhart (2002), das in der Folge näher erläutert wird, nimmt diesen Aspekt auf und bezieht die Fachlichkeit direkt in die Modellierung ein.

Modell von Terhart (2002)

Das mehrdimensionale normative Modell von Terhart (2002), das im Rahmen einer Expertise für die Kulturrministerkonferenz (KMK) entwickelt worden ist, umfasst Standards für die erste, universitäre Phase der Lehrerbildung sowie Standards für die zweite Phase im Studienseminar. Terhart (2002) unterscheidet darin zunächst Inhaltsstandards für die beiden Phasen der Ausbildung sowie eine Taxonomie von Kompetenzfacetten (Wissen, Reflektieren, Kommunizieren, Beurteilen und Können) (cf. Baumert & Kunter, 2006) und ergänzt das Modell mit einer dritten Dimension, welche die zeitliche, berufsbiographisch getaktete Perspektive der Kompetenzgenese abbildet. Für die erste, universitäre Studienphase sind, wie Abbildung 4 zeigt, wissensbasierte Standards beschrieben, welche die Bereiche Unterrichtsfach, Fachdidaktik, Erziehungswissenschaften und Schulpraxis abdecken. Diese Standards beschreiben fast ausnahmslos das Wissen zu bestimmten Inhalten aus den genannten Bereichen als zentrale Bestandteile der Kompetenz. Erst in den Standards im Studienseminar wird die Performanz in Form von Handlungsanforderungen aufgeführt. Darunter ist die Unterrichtsplanung explizit erwähnt. Ergänzend dazu sind auch weitere Standards beschrieben, die für die Planung ebenfalls bedeutsam sind (z.B. Einsatz eines breiten Methodenrepertoires, Einsatz neuerer Informationstechnologien, Beurteilung, Diagnose und Förderung). Aufgrund der Anlage dieses Modells kann angenommen werden, dass das Wissen (u.a. Fachwissen und fachdidaktisches Wissen) implizit als Konzept zur Beschreibung der Performanz aufgefasst wird (Vogelsang, 2014, S. 38).

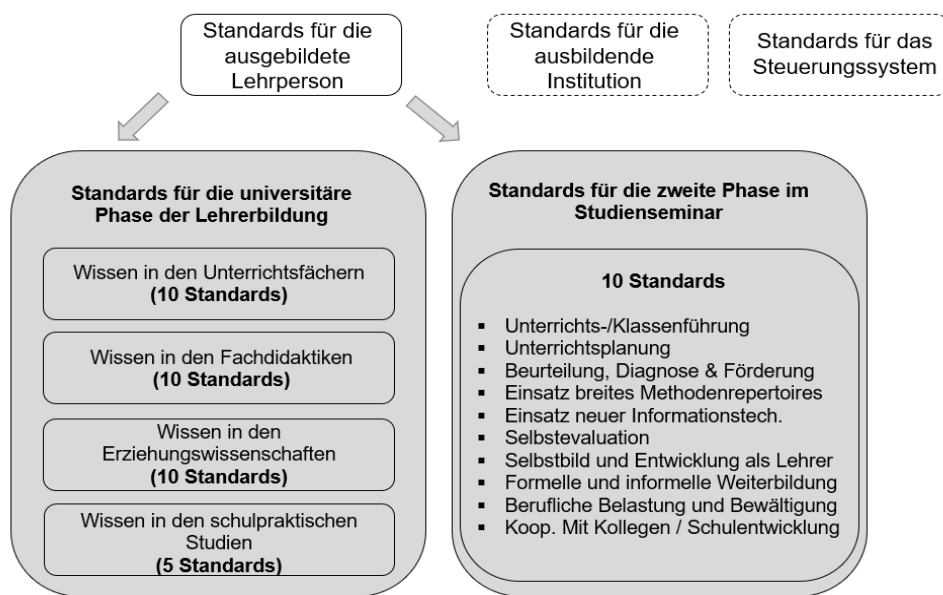


Abbildung 4: Standards der Lehrerbildung nach Terhart (2002)

Die Fachlichkeit spielt damit in diesem Modell eine Rolle. Terhart erachtet es als Notwendigkeit, fachspezifische Standards zu formulieren, da die erforderlichen Kompetenzen und somit deren zuverlässige Beurteilung fachspezifisch variieren bzw. fachspezifische Schwerpunkte aufweisen (zitiert in Frey, 2014, S. 730). Dieses Modell bildete die Grundlage für die Bildungsstandards der Lehrpersonenbildung in Deutschland. Diese werden nächstfolgend kurz dargestellt.

Bildungsstandards für die Lehrpersonenbildung in Deutschland (KMK, 2004)

Auf der Grundlage des Standardmodells von Terhart (2002) sind in Deutschland die Bildungsstandards für die Lehrpersonenbildung entwickelt worden (KMK, 2004). Diese bieten seit ihrem Beschluss vom 16.12.2004 einen Orientierungsrahmen für alle Phasen der Lehrpersonenbildung in Deutschland und beinhalten Zielvorgaben für die theoretische und für die praktische Ausbildung. Die Bildungsstandards umfassen mit dem Unterrichten, Erziehen, Beurteilen und Innovieren insgesamt vier Kompetenzbereiche, denen ihrerseits total 11 Kompetenzen jeweils für beide Ausbildungsphasen zugewiesen sind. Die Unterrichtsplanung ist im Kompetenzbereich Unterrichten als Kompetenz mit dem folgenden Wortlaut aufgeführt: „Lehrer planen Unterricht fach- und sachgerecht und führen ihn fachlich und sachlich korrekt durch“ (KMK, 2004, S. 7). Hierzu sind Standards sowohl für die theoretische wie auch für die praktische Ausbildung formuliert. Besonders hervorzuheben ist in diesem Zusammenhang ein Standard für die praktische Ausbildung, welcher die Verknüpfung fachlicher und fachdidaktischer Argumente zur Planung von Unterricht vorsieht.

Alle bisher beschriebenen standardorientierten Kompetenzmodelle, das HarMoS-Kompetenzmodell für den naturwissenschaftlichen Unterricht in der Schweiz eingeschlossen (cf. Kap. I, 2.1.3), zeichnen sich durch ihre Outputorientierung und damit durch ihre Performanznähe aus. Mit den beschriebenen Handlungsanforderungen bilden sie den Aspekt der Handlung und damit die Anwendung von Wissen stärker ab. In den Modellen von Oser (2001), Terhart (2002) und auch den Bildungsstandards Deutschlands (KMK, 2004) ist die Handlungsanforderung der Unterrichtsplanung implizit bzw. explizit enthalten. Blömeke (2006, S. 27) kritisiert in diesem Zusammenhang aber, dass diese Standards bei genauerer Betrachtung nur den erziehungswissenschaftlichen Teil der Lehrerbildung abdecken und dass Fachbezüge kaum vorkommen. Die Fachlichkeit, die gemäss Klieme-Expertise (Klieme et al., 2003, S. 24ff.) als erstes Qualitätsmerkmal für gute Bildungsstandards gilt und entsprechend gut ausgebildete Lehrpersonen verlangt, ist auch nur in den Modellen von Terhart (2002) und der KMK (2004) berücksichtigt und in letzterem für das Fach Biologie weiter präzisiert (KMK, 2005, cf. Kap. I, 2.1.2). Wilhelm (2007, S. 70) stellt hierzu fest, dass im Modell von Terhart (2002) weniger als 5% der Standards tatsächlich fachdidaktischer Natur sind. Standardorientierte Modelle sind gemäss diesen Ausführungen demnach stark erziehungswissenschaftlich ausgerichtet. Sie messen der Fachlichkeit weniger Bedeutung zu und müssen jeweils auf das entsprechende Fach übersetzt werden.

Professionswissensorientierte Kompetenzmodelle stellen demgegenüber die Fachlichkeit ins Zentrum. Hier zuordnen lassen sich u.a. Modelle von Shulman (1986/1987), Bromme (1997) sowie das im Rahmen des Projekts COACTIV entstandene Modell („Professionelle Kompetenz von Lehrkräften, kognitiv aktivierender Unterricht und die Entwicklung mathematischer Kompetenz“, cf. Brunner et al., 2006; Krauss et al., 2004; Kunter, Klusmann & Baumert, 2009). Das Modell der COACTIV-Studie (Baumert & Kunter, 2011a) bildet Ausgangspunkt oder Grundlage vieler naturwissenschaftsdidaktischer Forschungsprojekte. Es soll daher im nächsten Abschnitt vorgestellt werden.

II. Professionswissensorientierte Modelle

Das viel beachtete, nicht hierarchische Struktur- und Rahmenmodell der COACTIV-Studie beruht auf dem kompetenztheoretischen Ansatz (Baumert & Kunter, 2011a). Es wurde zur Erfassung des Professionswissens von Lehrpersonen und angehenden Lehrpersonen des Fachs Mathematik normativ entwickelt und empirisch geprüft. Im Vergleich zu den outputorientierteren Standardkompetenzmodellen entspricht es einem eher klassischen Modellansatz (Frey, 2014, S. 716), der sich stärker auf die eigentlichen Voraussetzungen des Handelns bezieht.

Das Modell basiert auf:

- dem Kompetenzverständnis von Weinert (2001a; cf. Kap. I, 2.4.1), gepaart mit den taxonomischen Ansätzen zum Professionswissen von Shulman (1986, 1987), den Kompetenzbereichen von Bromme (1997, 2001) und den Kernaussagen des National Board for Professional Teaching Standards (NBPTS, 2002) (cf. Baumert & Kunter, 2006, S. 481; Kunter et al., 2009, S. 154).
- Aspekten des Experten- und des Persönlichkeitsparadigmas.
- der Annahme, dass im Sinne der in Kapitel I, 2.4 beschriebenen *Wirkkette* Kompetenzausprägungen von Lehrpersonen Auswirkungen auf deren professionelles Verhalten haben und somit auch die Schüler- und Lehrerleistungen beeinflussen (Korneck, Kohlenberger, Oettinghaus, Kunter & Lamprecht, 2013).
- dem *Angebot-Nutzungsmodell* (Fend, 1998; Helmke, 2012; cf. Kap. I, 2.6.1), das Unterrichtsqualität als Ergebnis einer Ko-Konstruktion zwischen Lehrperson und Schülerin bzw. Schüler auffasst.
- auf Forschungstraditionen wie z.B. dem verständisvollen Lernen im Sinne kognitiv aktivierenden Unterrichts, dem Professionswissen in der Form von Expertise ohne Höchstleistungs- und Perfektionsanspruch sowie auch motivations- und organisations- bzw. gesundheitspsychologischen Grundlagen.

Das dem Modell zugrunde liegende Lehr- und Lernverständnis sowie wichtige Grundlagen für die Modellierung des Professionswissens sind in Tabelle 2 zusammengestellt.

Tabelle 2: COACTIV-Modell – Lernverständnis und Verständnis zum Professionswissen (cf. Baumert et al., 2011, S. 12ff.)

Verständnisvolles Lernen ...	Professionswissen ...
... ist ein aktiver, individueller Konstruktionsprozess	... ist <u>domänenspezifisch</u> , ausbildungs- und trainingsabhängig
... ist abhängig von individuellen kognitiven Voraussetzungen, aber vor allem vom bereichsspezifischen Vorwissen	... ist mit zunehmender Expertise besser vernetzt und hierarchisch organisiert
... erfolgt stets situiert und kontextuiert	... ist im Sinne von Fach- und Handlungswissen um Schlüsselkonzepte und eine begrenzte Zahl an Ereignisschemata (Skripts) organisiert
... wird durch motivationale und metakognitive Prozesse gesteuert	... integriert Kontexte und ist in Form von Handlungsrouninen oder Basisprozeduren automatisiert, aber gleichwohl an die spezifischen Bedingungen des Einzelfalls und des Kontextes adaptierbar
... wird durch kognitive Entlastungsmechanismen unterstützt (informationsreiche Wissensseinheiten; Automatisierung von Handlungsabläufen und Denkvorgängen)	

Die Zusammenstellung zeigt, dass sich das Modell der COACTIV-Studie stärker auf die personalen Voraussetzungen bezieht, die Lehrpersonen befähigen, ihren Beruf erfolgreich auszuüben (Baumert et al., 2011, S. 8). Dem Fachinhalt und seiner Bedeutung für das Denken, Wissen und Handeln von Lehrpersonen wird in diesem Modell speziell Rechnung getragen. Dabei wird die Erwerbbarkeit, Veränderbarkeit und Ausbildungsabhängigkeit professionellen Wissens besonders betont, insbesondere vor dem Hintergrund, dass die operative Gestalt erst durch die praktische Erfahrung erreicht wird (cf. Bromme, 2001; Neuweg, 2005).

Das Rahmenmodell professioneller Handlungskompetenz der COACTIV-Studie berücksichtigt neben dem aus der Expertiseforschung entwickelten Professionswissen (Bromme, 1997) auch motivational-selbstregulative Kompetenzfacetten und damit auch affektive Dispositionen im Sinne Weinerts (2001a). Dazu gehören *Überzeugungen und Werthaltungen*, *motivationale Orientierungen* und auch *selbstregulative Fähigkeiten* (Kunter et al., 2009, S. 155; Abb. 5).

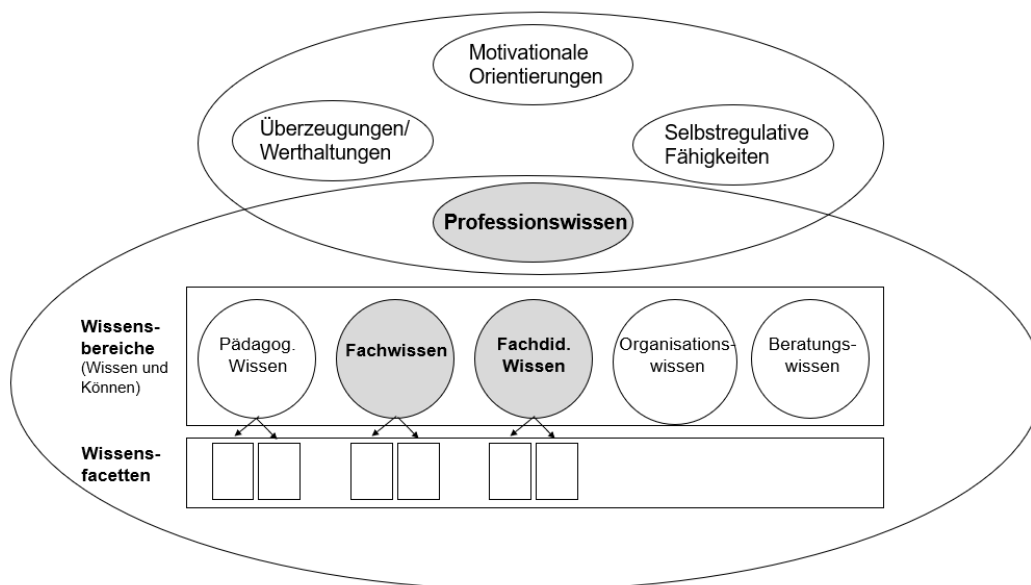


Abbildung 5: Modell professioneller Handlungskompetenz – Professionswissen (COACTIV; Baumert & Kunter, 2006, S. 482)

Die Kompetenzfacette *Überzeugungen und Werthaltungen* basiert auf einem Ordnungssystem, das systematische Wertbindungen, epistemologische Überzeugungen, subjektive Theorien über Lehren und Lernen sowie Zielsysteme für Curriculum und Unterricht unterscheidet (Baumert & Kunter, 2006, S. 497). Die COACTIV-Studie konzentriert sich auf die letzten drei Facetten, also Kompetenzbereiche, die handlungsrelevant für den Unterricht sind.

Die epistemologischen Überzeugungen bezüglich Struktur, Genese und Validierung von Wissensbeständen sowie die Überzeugungen bezüglich Lehren und Lernen von Mathematik werden unter dem Begriff der lerntheoretischen Überzeugungen zusammengefasst.

Motivationale Orientierungen sind für die psychische Dynamik des beruflichen Handelns verantwortlich (Baumert & Kunter, 2011a, S. 42). Die Bereitschaft, sich langfristig zu engagieren, hängt letztlich von der Motivation der Lehrperson ab (Kunter, 2011), die auch eine handlungssteuernde Funktion einnimmt (Brunner et al., 2006). Es liegen derzeit nur wenige empirische Erkenntnisse darüber vor, welche motivationalen Qualitäten den Unterrichtserfolg von Lehrpersonen tatsächlich vorhersagen lassen (Kunter, 2011, S. 261). Drei zentrale Bereiche, die sich auf Bereitschaft und Intentionen einer Person beziehen, sind aber Gegenstand diverser Untersuchungen: Die Berufswahlmotive, die Selbstwirksamkeitserwartungen sowie die intrinsischen Orientierungen bzw. der Enthusiasmus der Lehrpersonen. Die letzteren beiden bilden in der COACTIV-Studie den Kern der Kompetenzfacette motivationaler Orientierungen. Die Selbstwirksamkeitsüberzeugungen werden in Bezug auf den Unterricht und sonstige schulische Anforderungen, der Lehrerenthusiasmus als Personenmerkmal der Lehrpersonen konzeptualisiert (Kunter et al., 2008). Lehrerenthusiasmus wird – in Anlehnung an die erweiterte Erwartungs-Wert-Theorie (Wigfield & Eccles, 2000), die Theorie des individuellen Interessens (Krapp, 2000) und die Selbstbestimmungstheorie von (Deci & Ryan, 2000) – als den Grad des positiven emotionalen Erlebens während der Ausübung der Lehrertätigkeit aufgefasst (Baumert & Kunter, 2011a, S. 44).

Selbstregulative Fähigkeiten sind im Hinblick auf die hohen psychischen Anforderungen der Lehrertätigkeit, insbesondere den verantwortungsvollen Umgang mit den eigenen persönlichen Ressourcen erforderlich, um erfolgreich beruflich zu handeln (Klusmann, Kunter, Trautwein & Baumert, 2006). Dazu gehören Strategien zu dem Umgang mit und der Einstellung gegenüber der eigenen Arbeit. Die COACTIV-Studie geht dabei basierend auf Arbeiten anderer Forschungsgruppen von der Annahme aus, dass ein ausgewogenes Haushalten mit den eigenen Ressourcen, das heisst eine Balance aus Engagement und Distanzierungsfähigkeit, die beste Strategie sowohl für das eigene Befinden als auch für die Qualität des Unterrichts ist und definiert Selbstregulation entsprechend (Klusmann & Kunter, 2008).

Im Zentrum des kompetenztheoretischen und domänenspezifisch angelegten COACTIV-Modells stehen kognitive und affektive Dispositionen. Mit der hier beschriebenen Fokussierung wird das Handeln im Unterricht und damit die Performanz bewusst abgegrenzt. Damit kann das unterrichtliche Handeln gesondert erfasst und so in Beziehung zu den

kognitiven und affektiven Dispositionen gesetzt werden (Stender, 2014, S. 12). Das allgemeindidaktische Konzeptions- und Planungswissen ist als Facette generischen pädagogischen Wissens im Rahmenmodell enthalten (Baumert & Kunter, 2011a, S. 39), findet in diesem Modell jedoch auch nur unterschwellig Beachtung (Wernke & Zierer, 2017).

Die vorangehend beschriebenen standard- und professionswissensorientierten Modelle unterscheiden sich in ihrer Performanznähe, in der Berücksichtigung der Fachlichkeit wie auch hinsichtlich des Einbezugs der Unterrichtsplanung als eigentlichen Kompetenzbereich. Aufgrund des expliziten und im Fokus stehenden Fachbezugs berufen sich viele Studien der naturwissenschaftsdidaktischen Forschung auf das kompetenztheoretische Modell der COACTIV-Studie. Auch die vorliegende Studie stützt sich auf das professionswissensorientierte Strukturmodell professioneller Handlungskompetenz und die darin enthaltene Taxonomie professionellen Wissens. Die Frage nach der Topologie des Professionswissens, insbesondere des fachdidaktischen Wissens, das u.a. auch in Unterrichtsplanungen beobachtet werden kann, rückt vor diesem Hintergrund ins Zentrum.

Das nächstfolgende Kapitel widmet sich zum einen dem Professionswissen ganz allgemein, zum anderen aber auch dem fachdidaktischen Wissen zur Planung von Experimentalstunden mit Fokus auf der Erkenntnisgewinnung, das im Rahmen der vorliegenden Untersuchung in Unterrichtsplanungen und Planungsgesprächen von angehenden Lehrpersonen analysiert wird.

2.5 Professionelles Wissen von Lehrpersonen – kognitive Disposition professioneller Handlungskompetenz

In der Lehrpersonenbildungsforschung besteht heute weitgehend Einigkeit, dass lernwirksames Lehren und Unterrichten weniger auf Charaktereigenschaften und Talente zurückzuführen ist, als – im Sinne des «Expertenparadigmas» (cf. Kap. I, 2.3.3) – auf das Wissen und Können von Lehrpersonen (cf. Besser & Krauss, 2009; Bromme, 2008; Jüttner & Neuhaus, 2013). Insbesondere das Professionswissen wird für eine lernförderliche Unterrichtsgestaltung als bedeutsame Einflussvariable eingestuft (Abell, 2007; Baumert et al., 2010; Blömeke et al., 2009; Blömeke, Kaiser & Lehmann, 2008; Borowski et al., 2010; Gess-Newsome, 1999; Grossman, 1990; Kirschner et al., 2017; Kunter et al., 2009; Schmelzing, Wüsten, Sandmann & Neuhaus, 2008). Weniger Übereinstimmung besteht hingegen in Bezug auf die strukturbildenden Dimensionen des Wissens, die verschiedenen Wissensfacetten und -typen sowie die Entstehung und mentale Repräsentation professionellen Wissens und Könnens (Abell, 2007; Besser & Krauss, 2009; Bromme, 1997; Fenstermacher, 1994; Neuweg, 2014; Shulman, 1987, 1986; Stern et al., 2009). Baumert & Kunter (2011a, S. 33)

führen dies auf mangelnde empirische Evidenz und Unterschiedlichkeit in den theoretischen Perspektiven zurück. Zur Konzeptionalisierung und Operationalisierung des Professionswissens von Lehrpersonen sind daher in der Literatur verschiedene Auffassungen und Ansätze zu finden.

2.5.1 Taxonomie des Professionswissens von Lehrpersonen – zentrale Wissensdimensionen

Zur Taxonomie des Professionswissens von Lehrpersonen gibt es in der Literatur eine Vielzahl an Vorschlägen, die sich auf Shulman (1986) zurückführen lassen (Gramzow, 2015, S. 31). Shulman (1986) identifizierte in seiner Topologie des Professionswissens als einer der ersten drei fachbezogene Dimensionen des professionellen Wissens von Lehrpersonen, denen er besondere Bedeutung zugeschrieben hat, ohne dabei dem allgemeinen pädagogischen Wissen (*general pedagogical knowledge*) seine Wichtigkeit abzusprechen. Es sind dies das Fachwissen (*subject matter content knowledge*), das fachdidaktische Wissen (*pedagogical content knowledge*) und das Wissen über das Fachcurriculum (*curriculum knowledge*). Diese insgesamt vier Dimensionen erweiterte er später mit den Dimensionen Psychologie des Lerners (*knowledge of learners*), Organisationswissen (*knowledge of educational context*) sowie erziehungsphilosophisches, bildungstheoretisches und bildungshistorisches Wissen (Shulman, 1987). Insbesondere drei dieser sieben Dimensionen des Professionswissens von Lehrpersonen haben sich in der Lehrpersonenbildungsforschung etabliert (Bromme, 1997). Die Rede ist von: *content knowledge* (CK, Fachwissen), *pedagogical content knowledge* (PCK, fachdidaktisches Wissen) und *pedagogical knowledge* (PK, pädagogisches Wissen). Diese drei Wissensdimensionen werden seither als wesentliche Elemente professioneller Kompetenz angesehen und sowohl im deutschsprachigen als auch im angloamerikanischen Raum als besonders relevant für die Gestaltung eines lernförderlichen Unterrichts angenommen (Abell, 2007; Baumert et al., 2010; Blömeke et al., 2009; Blömeke, Kaiser, et al., 2008; Depaepe, Verschaffel & Kelchtermans, 2013; Gess-Newsome, 1999; Grossman, 1990; Kunter et al., 2009). Sie entsprechen auch den drei Kernbereichen der Lehrpersonenbildung in Deutschland und in der Schweiz (Blömeke, Felbrich, et al., 2008; Fischler, 2008) und bilden eine wichtige Grundlage aktueller empirischer Forschungsprojekte, auch in der naturwissenschaftlichen Lehr- und Lernforschung (e.g. von Aufschnaiter & Blömeke, 2010; Borowski et al., 2010; Schmelzing, Wüsten, Sandmann & Neuhaus, 2010; Tepner et al., 2012). Im angloamerikanischen Raum werden diese Wissensdimensionen in manchen Modellen noch durch eine weitere, vierte Dimension, bezeichnet mit *contextual knowledge*, ergänzt (Abell, 2007; Grossman, 1990, S. 9; cf. Magnusson, Krajcik & Borko, 1999). Damit soll dem spezifischen Wissen zu bestimmten

Schulsettings oder auch zu unterschiedlichen Schülerinnen und Schülern explizit Rechnung getragen werden.

Gemäss Tillema (2009) bieten Unterrichtsplanungen einen geeigneten Zugang, um Einblick in die Kompetenzen der Studierenden zu erhalten. Das in diesem Zusammenhang beobachtbare fachdidaktische Wissen wird von Shulman als wichtige Grundlage für die Ausübung der Lehrprofession angesehen (cf. Kind, 2009, S. 173). Die theoretische Konzeptualisierung dieser für die Profession, insbesondere aber auch für das Unterrichten eines Faches oder Unterrichtsthemas als exklusiv erachteten Wissensdimension, soll im nächsten Abschnitt näher vorgestellt werden. Auf das Fachwissen als weitere Dimension wird erst im zweiten Teil dieser Arbeit näher eingetreten (cf. Kap. II, 1.1.1).

2.5.2 Das fachdidaktische Wissen bzw. *pedagogical content knowledge* – Formen der Modellierung

Mit dem fachdidaktischen Wissen identifiziert Shulman eine spezielle Wissensart, durch die sich Lehrpersonen von Fachexperten unterscheiden und über die sie u.a. verfügen sollten, um erfolgreich unterrichten zu können (Shulman, 1986). Er unterstreicht dies mit der Aussage: „Pedagogical content knowledge is that special amalgam of content and pedagogy, that is uniquely the province of teachers, their own special form of professional understanding“ (Shulman, 1987, S. 8).

Das fachdidaktische Wissen erachtet er demzufolge als Kombination und Integration von fachspezifischem und pädagogischem Wissen, das Lehrpersonen dazu befähigt, Fachinhalte gemäss Interessen und Fähigkeiten von Lernenden in fruchtbare Lerngelegenheiten zu übersetzen (Shulman, 1987, S. 8). Damit ist das fachdidaktische Wissen mehr als die Summe seiner einzelnen Wissensdimensionen. Aufgrund der engen Verbindung zum Fachwissen ist es ausserdem spezifisch für ein Fachgebiet oder einen Themenbereich und damit auch für den entsprechenden Unterricht (Abell, 2008; Smith & Banilower, 2015; cf. van Driel, Verloop & de Vos, 1998). Diese Wissensdimension stellt einen wichtigen Teil professioneller Handlungskompetenz dar, im Sinne des kompetenztheoretischen Ansatzes von Baumert & Kunter (2006) (Kapitel I, 2.2.2). Shulmans Konzeptualisierung von PCK ist aber auch kritisiert worden. Kritikpunkte waren (1) der fehlende theoretische und empirische Hintergrund, (2) die zu statische Auffassung dieser Wissensdimension, (3) die Unklarheit, ob es sich um eine vom Fachwissen abgrenzbare Wissensdimension handelt, (4) die zu enge Konzeptualisierung und (5) der normative Charakter dieser Wissensdimension (cf. Depaepe et al., 2013; Kind, 2009). Als Antwort auf die Kritik sind neue Definitionen und Modellierungen dieser Wissensdimension entstanden. Dabei fehlen einheitliche

Definitionen und Modellierungen nach wie vor (cf. für eine Übersicht Depaepe et al., 2013). Die nächstfolgenden Abschnitte beleuchten diese Sachlage.

I. Fachdidaktisches Wissen (FDW) und *pedagogical content knowledge* (PCK) – Definition im Vergleich

Das fachdidaktische Wissen (FDW)³ als Wissensdimension wird in der Literatur häufig synonym mit dem englischen Begriff *pedagogical content knowledge* (PCK) verwendet. Gemäss Fischler (2011) sind jedoch die den Begriffen zugrundeliegenden Konzepte nicht ganz deckungsgleich. Die zu verzeichnenden Unterschiede sind auf kulturelle Traditionen und sich unterscheidende Rahmenkonzeptionen zurückzuführen (Gramzow, 2015, S. 73). Während der deutsche Begriff das Bildungsverständnis und die damit verbundenen Anforderungen mit einschliesst, liegt der Fokus des englischen Begriffs stärker auf der Vermittlung curricularer Inhalte und stellt so ein engeres, aber auch forschungsaffineres Konzept im Sinne der Wirkkette schulischer Bildung dar (Kansanen, 2009). Dabei ist im Konzept von PCK die Überprüfung der Relevanz und Transformation des zu vermittelnden Fachinhaltes in einen für Schülerinnen und Schüler bedeutsamen und stufenadäquaten Unterrichtsinhalt nicht enthalten (Bromme, 1995). Dies wird gemäss Fensham (2001) aber durch das Curriculum abgedeckt. Das Konzept des FDW ist damit weitreichender und schliesst PCK mit ein (Gramzow, Riese & Reinhold, 2013). Es wird als reine kognitive Disposition beschrieben. PCK wird demgegenüber sowohl als kognitive Ressource, als auch als „*knowledge in action*“ handlungsnäher aufgefasst (Gramzow, 2015, S. 89). Kind unterstreicht dies mit der Aussage „PCK is a concept, that represents the knowledge teachers use in the process of teaching“ (Kind, 2009, S. 170).

Für die Untersuchung des Professionswissens und der Wirksamkeit der Lehrpersonenbildung liefern beide vorangehend beschriebenen Konzepte fachdidaktischen Wissens⁴ einen weitgehend akzeptierten theoretischen Rahmen (cf. Abell, 2008). Jedoch liegen weder für das FDW noch für PCK einheitliche Konstrukte vor (Abell, 2007). Die Modellierung beider Wissensdimensionen ist in der Expertiseforschung seit Beginn der neunziger Jahre Gegenstand kontroverser Diskussionen. In der Literatur lassen sich hierzu verschiedene, zum Teil konträre, theoretische Positionen und entsprechende Modelle finden. Zum einen sind dies Modelle, die das FDW sowie PCK als Disposition im Sinne von „*knowledge needed*

³ Diese Abkürzung wird zur besseren Unterscheidung und in Anlehnung an Gramzow (2015) verwendet.

⁴ Der Terminus «fachdidaktisches Wissen» wird verwendet, wenn die Wissensdimension als solche angesprochen ist und keine Unterscheidung zwischen den beiden Konzepten FDW & PCK vorgenommen wird.

for teaching a specific subject“ (Krauss, Brunner et al., 2008, S. 716) beschreiben. Dazu gehören *transformative* und *integrative* Modelle (cf. Gess-Newsome, 1999).

Zum anderen sind es aber auch Modelle dynamischer Natur, die PCK handlungs- bzw. performanznäher auffassen, wie z.B. das neuere *Konsensmodell* (Shulman, 2015). Diese Modelle sollen in der erwähnten Reihenfolge näher erläutert werden.

II. Transformative und integrative Modelle fachdidaktischen Wissens (FDW bzw. PCK) – eine Frage der Dimension

Transformative Modelle fassen das fachdidaktische Wissen als eine eigenständige Wissensdimension auf, die aus der Transformation des Fachwissens, pädagogischen Wissens und Kontextwissens entsteht (Abb. 6). Hierzu lassen sich alle Modellierungen, die auf der Grundlage von Shulmans Modell erweitert oder angepasst worden sind, zuordnen (Abell, 2008; Bromme, 1997; Carlsen, 1999; Grossman, 1990; Krauss, Neubrand, et al., 2008; Magnusson, Krajcik & Borko, 1999; Park & Chen, 2012; Shulman, 2015; Tamir, 1988). Auch das im deutschsprachigen Raum gängige Modell der professionellen Handlungskompetenz aus der COACTIV-Studie (Kap. I, 2.4.2) fasst das fachdidaktische Wissen als eigenständige, jedoch aber nicht überschneidungsfreie Dimension auf (Gramzow, et al., 2013).

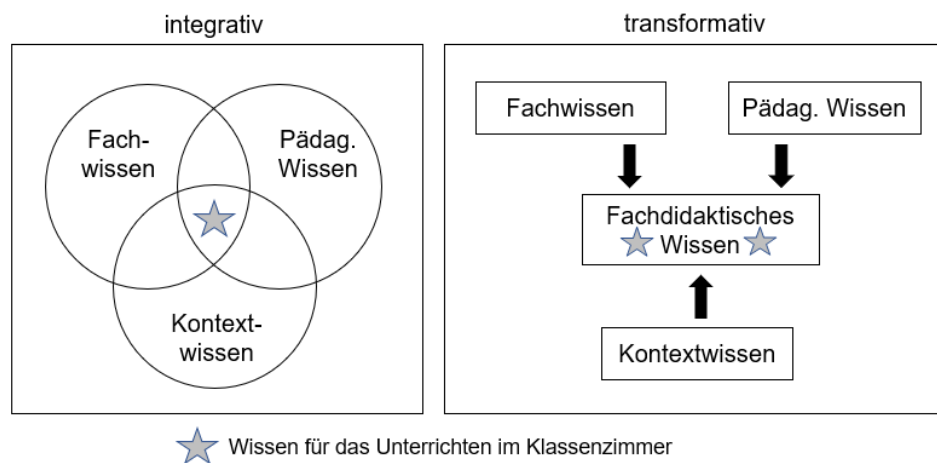


Abbildung 6: Modelltypen zur Rolle des fachdidaktischen Wissens (Gess-Newsome 1999, S. 12)

Integrative Modelle interpretieren demgegenüber PCK als Schnittmenge von Fachwissen, pädagogischem Wissen und Kontextwissen (e.g. Cochran, DeRuiter & King, 1993; Hashweh, 2005; Abb. 6). PCK schliesst in diesen Modellen andere Wissensbereiche mit ein und ist damit empirisch nicht von diesen trennbar. Entsprechend unterscheiden sich diese Modelle grundlegend von Shulmans Modell.

Transformative wie auch *integrative* Modelle sehen ungeachtet ihrer unterschiedlichen Modellierungen einen direkten Zusammenhang zwischen dem fachdidaktischen Wissen und der Planung bzw. dem unterrichtlichen Handeln (Fernandez, 2014, S. 84). *Integrative* Modelle umreißen dabei aber ein eher allgemeineres Bild von diesem Wissen und den Fähigkeiten, die Lehrpersonen benötigen, während *transformative* Modelle vor allem das fach- und themenspezifische fachdidaktische Wissen fokussieren (Kind, 2009; cf. van Dijk & Kattmann, 2010; van Driel et al., 1998).

Wie bereits erwähnt, wurde Shulmans Modell aufgrund seines fehlenden theoretischen und empirischen Hintergrunds auch kritisiert. Oser, Heinzer & Salzmann (2010) weisen zunehmend darauf hin, dass sich mit diesem Modell keine umfassenden Kompetenzmessungen vollziehen lassen, da der Fokus ausschliesslich auf dem Wissen liegt. Auch kritisieren sie Diagnoseinstrumente, die Kompetenzen unabhängig von situativen Kontexten erfassen. Dies unterstreichen sie mit der Aussage: “Das eigentlich Zentrale des Lehrerberufs als Profession – nämlich die Situativität, Authentizität, Komplexität und die Kontextgebundenheit des unterrichtlichen Handelns – wird ... [damit] vernachlässigt“ (Oser et al., 2010). Hasse et al. (2014) heben in ihrer Arbeit hervor, dass Wissen und Kompetenzen zwei verschiedene Konstrukte darstellen. Während Kompetenzen den Fokus auf die Problemlösefähigkeiten und -fertigkeiten legen (Weinert, 2001b), stellt das Wissen eine notwendige und wichtige Grundlage für die Erlangung einer Kompetenz dar (Weinert, 1999). Vogelsang und Reinhold (2013a) betonen, dass ohne die Betrachtung der Handlungsebene die Handlungsrelevanz von erfasstem Professionswissen nicht eindeutig begründbar ist. Zur Einschätzung des Einflusses auf die Lernenden ist daher ihrer Meinung nach wirkliches Unterrichtshandeln notwendig.

Shulman (2015) selbst weist in neuerer Zeit auf Schwächen seines ursprünglichen transformativen Konstrukts von PCK hin und erwähnt dazu u.a. folgende Kritikpunkte:

- Das Modell ist zu statisch.
- Die kognitive Dimension ist überbewertet und der Bezug zur Performanz im Klassenzimmer fehlt.
- Es fehlt im Modell die affektive Dimension.
- Es fehlt auch der Kontextbezug.
- Auch fehlt ein Bezug zu den persönlichen Vorstellungen und Bildungszielen der Lehrperson.

Diese Kritikpunkte decken sich mit den zuvor erwähnten Ausführungen. Abell (2008, S. 1407) fügt hier noch ergänzend an, dass PCK mit den Aufgaben und Problemstellungen der Praxis eng verbunden sein sollte und dabei der thematische Bezug von zentraler Bedeutung ist. Vor diesem Hintergrund ist im Rahmen eines Workshops „PCK Summit“ ein

Konsensmodell zu PCK entwickelt worden, das diese Kritikpunkte aufnimmt und entsprechende Anpassungen vorgenommen hat (Berry, Friedrichsen & Loughran, 2015). Dieses Konsensmodell soll im nächsten Abschnitt näher vorgestellt werden.

III. Konsensmodell – ein dynamisches Modell von PCK

Das in Abbildung 7 dargestellte Konsensmodell ist ein Versuch, die Dynamik der unterrichtlichen Tätigkeit von Lehrpersonen im Sinne der Wirkkette möglichst ganzheitlich abzubilden. Entsprechend werden die für die Praxis erforderlichen Wissensdimensionen, das im Unterricht beobachtete Wissen sowie auch die Leistungen der Schülerinnen und Schüler im Modell berücksichtigt. Das Modell unterscheidet zwischen zwei kontextfreien Wissensdimensionen sowie dem kontextspezifischen PCK, das sich erst in unterrichtsnahen Situationen zeigt.

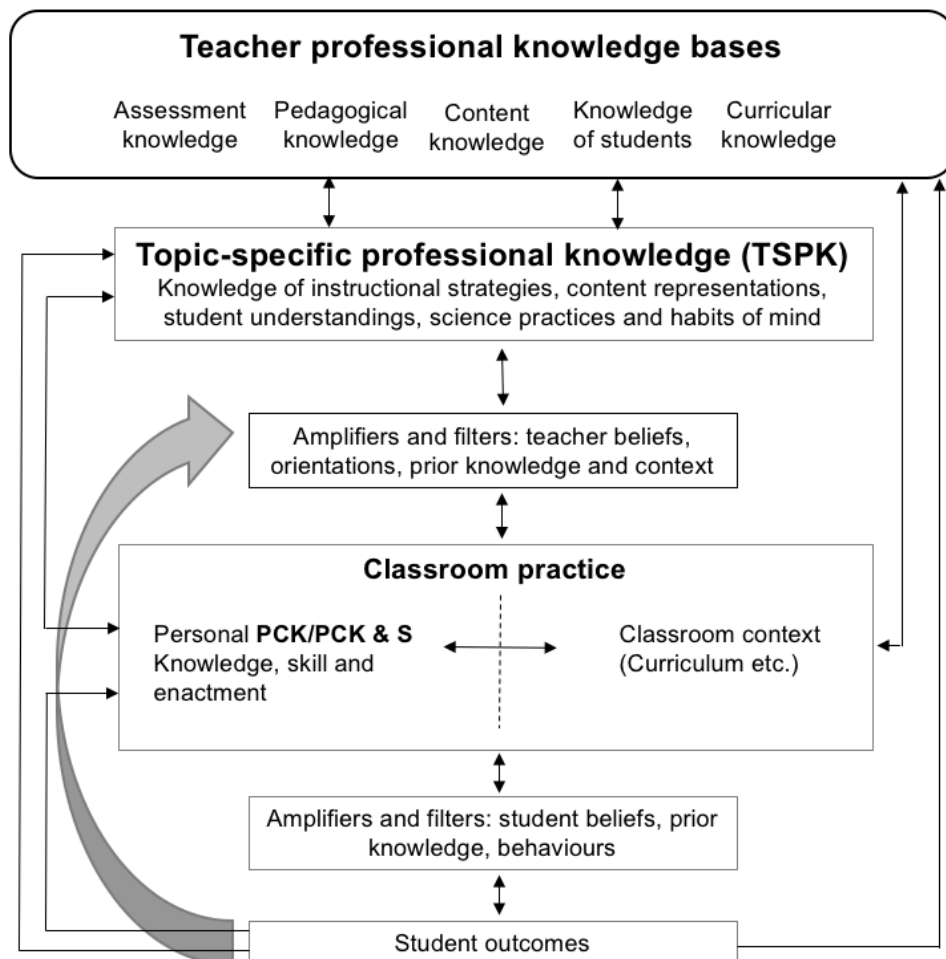


Abbildung 7: Das Konsensmodell (Gess-Newsome, 2015, S. 31)

Die ersten beiden Wissensdimensionen werden als kognitive Dispositionen im Sinne von Handlungsvoraussetzungen aufgefasst. Sie umfassen das allgemeine sowie das themen-

spezifische Professionswissen (TSPK). TSPK ist dabei normativ, auf der Grundlage empirischer Ergebnisse und Expertise festgelegt worden und beinhaltet, was eine Lehrperson für die Gestaltung eines qualitätvollen Unterrichts wissen muss. PCK wird demgegenüber als eine eigene Wissensdimension personen-, themen- und kontextspezifisch modelliert. Dabei wird es zum einen als kognitive, kontextspezifische Disposition im Sinne von Wissen oder *knowledge base* (PCK) beschrieben, zum anderen als PCK, das sich in der Handlung zeigt, also als *PCK and skill* (PCK&S).

Gess-Newsome (2015, S. 36) umschreibt diese beiden Konstrukte personenspezifischen PCKs wie folgt:

- „Personal PCK is the knowledge of, reasoning behind, and planning for teaching a particular topic in a particular way for a particular purpose to particular students for enhanced student outcomes (Reflection on Action, Explicit).”
- “Personal PCK&S is the act of teaching a particular topic in a particular way for a particular purpose to particular students for enhanced student outcomes (Reflection in Action, Tacit or Explicit)”.

PCK stellt im ersten Fall eine Anwendung von Wissen in Bezug auf die Unterrichtskonzeption dar. Diese lässt sich in Unterrichtsplanungen sowie auch mit Hilfe von Think-Aloud-Protokollen erfassen, in denen Entscheidungen zu gewählten Instruktionsstrategien begründet werden. PCK entspricht damit der *Reflection on Action* (Schön, 1983), ist zielgerichtet und soweit beobachtbar, explizit und sichtbar. Im Rahmen des Planungshandelns ermöglicht PCK damit aufgrund der sichtbaren Qualität dieses Wissens einen gewissen Einblick in das Können von (angehenden) Lehrpersonen. Im zweiten Fall bildet PCK einen Teil des Konstruktes PCK&S und findet seine Anwendung direkt im Unterricht, in welchem die Lehrperson die Planung umsetzt und unter Einbezug der Schülerinnen und Schüler die Instruktionsstrategien schnell anpassen muss. Zur Erfassung des angewendeten Wissens und der Gründe für erfolgte Entscheidungen während des Unterrichtens ist hier *Reflection in Action* beispielsweise in Form von Think-Aloud-Protokollen zu videografierten Unterrichtssequenzen möglich. Das zu erfassende Wissen ist dabei dynamischer, manchmal expliziter, oft aber auch impliziter Natur. Im Zentrum steht hier das Können der Lehrperson, das sich im tatsächlichen Unterricht zeigt und die Schülerleistungen entsprechend beeinflusst.

Gess-Newsome (2015, S. 40) erwähnt, dass aufgrund der durch das Konsensmodell erfolgten weiteren Klärung, die von ihr vorgeschlagene Einteilung in transformative und integrative Modelle nicht mehr hilfreich ist. Im neuen Modell wird explizit zwischen dem themenspezifischen fachdidaktischen Professionswissen (TSPK) und dem themen- und kontextspezifischen PCK unterschieden. Ersteres beruht auf theoretischen Konzepten, wird

normativ festgelegt und ist statisch. Letzteres ist personenspezifisch, zeigt sich erst in unterrichtsnahen Handlungen oder im Unterricht und ist dynamisch (cf. Nilsson & Vikström, 2015). Damit wird PCK in Form von zwei eigenständigen Wissensdimensionen beschrieben und weiter ausgeschärft, eine Unterscheidung, die schon Bromme 1995 gefordert hat (cf. Duit et al., 2012). Smith und Banilower (2015, S. 90), die sich mit der Assessmentfrage von TSPK und PCK befassen, erwähnen: „... the importance of being very clear about what „counts as PCK is paramount“. Sie unterscheiden zwischen normativ festgelegtem, themenbezogenem *collective PCK*, was im Modell dem *topic-specific professional knowledge* (TSPK) entspricht und dem *personal PCK*. Dabei vertreten sie die Ansicht, dass das normativ festgelegte *collective PCK* in der gleichen Weise wie auch das Fachwissen auf seine Richtigkeit hin geprüft werden kann und sich das individuelle, *personal PCK* erst in der Handlungssituation zeigt. Beide Formen von PCK lassen sich gemäss diesen Autoren aber nur in Form einer latenten Variablen erfassen (Smith & Banilower, 2015, S. 92).

Dieses hier beschriebene dynamische und rekursive Konsensmodell (Gess-Newsome, 2015) weist einige Parallelen auf zum prozessorientierten Kompetenzmodell von Blömeke, Gustafsson und Shavelson (2015, S. 9; cf. Kap. I, 2.4.1). Im Sinne einer Synopse lassen sich diese beiden Modelle in gewisser Weise zusammenführen (Abb. 8).

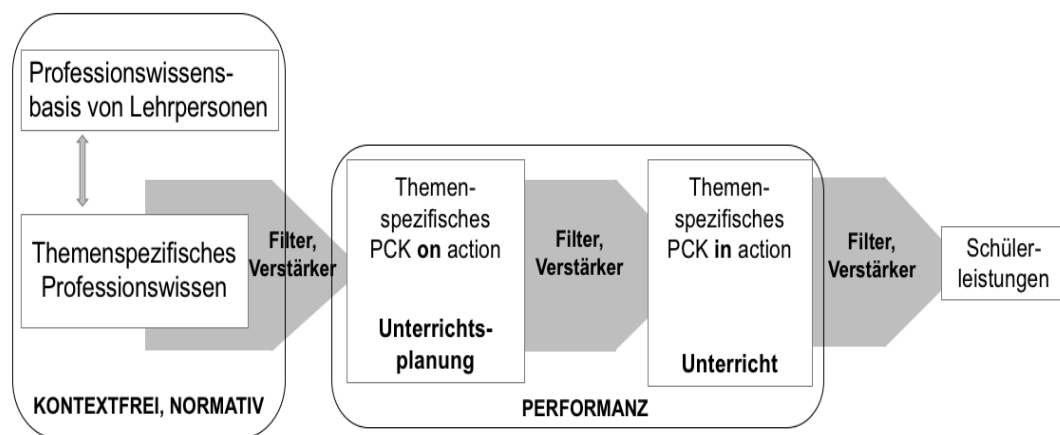


Abbildung 8: Synopse aus prozessorientiertem Kompetenzmodell und Konsensmodell (Blömeke et al., 2015; Gess-Newsome, 2015)

In beiden Modellen werden die Dispositionen einer Lehrperson mit ihrem beobachtbaren Verhalten, der Performanz, in Verbindung gebracht. Ausserdem fungieren personen- und kontextspezifische Filter bzw. Verstärker als Mediatoren und beeinflussen die Wahrnehmung von lernrelevanten Situationen, deren Analyse und Interpretation und die anschließende Entscheidung über Handlungsoptionen. Allerdings unterscheidet sich das Konsensmodell vom prozessorientierten Kompetenzmodell dahingehend, dass letzteres die

Wissensdimensionen im Hinblick auf PCK und andere Wissensdimensionen ausgeschärft hat und die Leistungen der Schülerinnen und Schüler im Sinne der Wirkkette auch im Modell enthalten sind.

IV. FDW und PCK – Modellierungen aus Sicht der naturwissenschaftsdidaktischen Forschung

Die naturwissenschaftsdidaktische Forschung im deutschsprachigen Raum orientiert sich am *transformativen* Modell von Shulman (1986). Entsprechend wird das FDW im Sinne gängiger Modelle professioneller Handlungskompetenz (e.g. Grossschedl, Harms, et al., 2015; Kirschner et al., 2017; Riese, 2009; Schmelzing, 2010; Tepner et al., 2012) als eine eigene Wissensdimension neben dem Fachwissen und dem pädagogischen Wissen verstanden. Alle drei Wissensdimensionen (Fachwissen, fachdidaktisches Wissen und pädagogisches Wissen) werden als notwendig für das Handeln im Unterricht erachtet (Baumert & Kunter, 2006). Ausserdem wird das FDW als rein kognitive Disposition beschrieben. Auch die vorliegende Studie interpretiert das fachdidaktische Wissen als eigene, aus der Transformation verschiedener Wissensfacetten entstandene themenspezifische Wissensdimension. Aufgrund der Analyse von schriftlichen Unterrichtsplanungen und videografierten Planungsgesprächen liegt der Fokus im Sinne von Gess-Newsome (2015) auf *PCK on action*. Der Untersuchung wird aber ein Modell zugrunde gelegt, das auf der «Didaktischen Rekonstruktion» basiert (Kattmann, Duit, Gropengiesser & Komorek, 1997; cf. Kap. I, 2.8.4). Entsprechend handelt es sich um ein etwas weiter gefasstes Verständnis von PCK – annäherungsweise im Sinne des fachdidaktischen Wissens (FDW).

Nicht nur die Modellierung fachdidaktischen Wissens als Wissensdimension erweist sich als grosse Herausforderung. Auch bei der Strukturierung innerhalb des Konstrukts liegen verschiedene Auffassungen vor, sei es in Bezug auf die Anzahl der Teildimensionen dieser Wissensdimension, sei es hinsichtlich der Frage, welche Kenntniskategorien bzw. Wissensfacetten tatsächlich zum FDW bzw. zu PCK gezählt werden (Riese & Reinhold, 2012a). Kirschner, Borowski, Fischer, Gess-Newsome und von Aufschnaiter (2016) führen dies auf unterschiedliche Forschungsziele zurück, was die Vergleichbarkeit der verschiedenen Konzepte erschwert. Der nächste Abschnitt geht auf diese Problematik ein.

V. Teildimensionen und Wissensfacetten von FDW und PCK – Sichtweise der naturwissenschaftsdidaktischen Forschung

Für die empirische Erfassung fachdidaktischen Wissens sind aufgrund unterschiedlicher Forschungsziele und damit auch unterschiedlicher Anforderungen eine Vielzahl an Konstrukten entwickelt worden. Diese unterscheiden sich zum einen in ihren Teildimensionen, zum anderen in ihren Wissensfacetten. In Bezug auf die Teildimensionen sind in der

Forschungsliteratur ein- bis dreidimensionale Modelle für das FDW (siehe Übersicht in Gramzow, et al., 2013) und meist eindimensionale Modelle für PCK vorzufinden. Konzeptualisierungen zum FDW sind damit aufgrund mehrerer Teildimensionen komplexer modelliert. Dreidimensionale Modelle zu FDW, zu denen u.a. Konzeptionen von Schmelzing (2010), ProwiN (Tepner et al., 2012) und KiL (Kröger, Neumann & Petersen, 2013) gehören, sind gekennzeichnet durch die Teildimensionen «Inhaltsbereiche oder Inhalte» (Fachliche Themen), «Wissensfacetten» im Sinne von Kenntniskategorien fachdidaktischen Wissens und «Wissensarten bzw. Wissenskomponenten». Die Konstrukte aus den naturwissenschaftsdidaktischen Projekten ProwiN und KiL sind ausserdem explizit fächerübergreifend angelegt. Ein zweidimensionales Modell, wie das Modell von Gramzow et al. (2013), unterscheidet im Vergleich dazu zwischen den Teildimensionen «Inhaltsbereiche» und «Wissensfacetten». Eindimensionale Modelle, wie z.B. Modelle von Riese (2009) und Baumert und Kunter (2011a), konzeptualisieren das FDW nur aufgrund der «Wissensfacetten». Diese werden bei Riese (2009) mit Bezug zu den Wissensarten (deklarativ – prozedural) auf einem Kontinuum verortet. Im Modell von Baumert und Kunter (2011a) werden die Wissensarten bzw. -typen definiert, für die Konzeptualisierung FDW jedoch nicht herangezogen. Beide Modelle beinhalten keine Teildimension mit fachlichem Fokus (Inhaltsbereiche), was bei Riese (2009) z.B. darauf zurückzuführen ist, dass nur ein einziger fachlicher Inhaltsbereich, nämlich die Mechanik, im Fokus steht.

Die auf der Grundlage von Shulmans Modell entwickelten Modellierungen von PCK sind demgegenüber mehrheitlich eindimensional und beinhalten nur die Teildimension der «Wissensfacetten». Dazu gehören u.a. Modelle von Park und Oliver (2008) oder auch Magnusson et al. (1999). Diese, hauptsächlich aus dem englischsprachigen Raum stammenden Modelle, unterscheiden sich allerdings im Hinblick auf ihre Wissensfacetten bzw. auf deren spezifische Kategorien von Kenntnissen z.T. erheblich. Erkenntnisse dazu wurden in Übersichtstabellen zusammengetragen (e.g. van Driel, Verloop & de Vos, 1998; Lee & Luft, 2008; Park & Oliver, 2008; cf. auch Depaepe et al., 2013). In neueren Arbeiten aus dem deutschsprachigen Raum sind diese Übersichtsdarstellungen zu den Wissensfacetten mit entsprechenden Konzeptionen des FDW ergänzt worden (Cauet, 2016; Kirschner et al., 2016; Schmelzing, 2010). Auch in diesen Zusammenstellungen wird deutlich, dass diese Teildimension je nach Konzeptualisierung eine unterschiedliche Anzahl an Wissensfacetten aufweist. Eine genaue Analyse ausgewählter Studien zum FDW bzw. zu PCK in den naturwissenschaftlichen Fächern, welche die Übersichtstabelle 3 auf der folgenden Seite zeigt, ergibt ein ähnliches Bild.

Tabelle 3: Modellierungen fachdidaktischen Wissens bzw. von PCK (mehrheitlich naturwissenschafts-
didaktische Forschungsprojekte, cf. Schmelzing, 2010, Kirschner et al., 2016)

Autoren	Facetten von fachdidaktischem Wissen/ PCK								
	Art der Studie	Fachinhalt	Pädagogik	Kontext	Schülerverständnis	Instruktions- strategien	Curriculum	Ziel des gelehrt en Fachinhalts	Leistungserfassung & -bewertung
Kirschner et al. (2016) (ProwiN)	m	–	–	–	O	O	–	–	–
Jüttner et al. (2013) (ProwiN)	m	–	–	–	O	O	–	–	–
Tepner et al. (2012) (ProwiN)	m	–	–	–	O	O	–	–	–
Grossschedl et al. (2015) (KiL)	m	–	–	–	O	O	O	–	O
Schmelzing et al. (2013)	m	–	–	–	O	O	–	–	–
Gramzow (2015) (ProfiL-P)	m	–	–	(O)	O	O	(O)	(O)	(O)
Riese (2009)	m	–	–	–	O	O	O	O	O
Brovelli et al. (2013)	m	–	–	–	O	O	–	–	–
Kotzebue L. & Nerdel C. (2015)	m	–	–	–	O	O	–	–	–
Stender, Brückmann, Neumann (2017)	m	–	–	–	O	O	O	–	O
Smith & Banilower (2015)	m	–	–	–	O	O	–	–	O
Ergönenç et al. (2014) (QuIP)	m	–	–	–	O	O	O	–	–
Gess-Newsome et al. (2017)	b	O	O	O	O	O	–	–	–
Alonzo & Kim (2016)	b	–	–	–	O	O	–	–	–
Brown et al. (2013)	b	–	–	–	O	O	–	O	–
Park & Suh (2015)	b	–	–	–	O	O	O	O	O
Loughran et al. (2012)	b	O	O	O	O	O	–	O	–
van Dijk & Kattmann (2010)	b	O	–	–	O	O	–	–	–
Park & Oliver (2008)	b	–	–	–	O	O	O	O	O
Hashweh (2005)	b	O	O	O	O	O	O	O	O
van der Valk & Broekman (1999)	b	–	–	–	O	O	–	–	–
Magnusson et al. (1999)	b	–	–	–	O	O	O	O	O
van Driel et al. (1998)	b	–	–	–	O	O	–	–	–
Fernandez-Balbao & Stiehl (1995)	b	O	–	O	O	O	–	O	–
Tamir (1988)	b	–	–	–	O	O	O	–	O
Shulman (1987)	b	–	–	–	O	O	–	–	–

Legende: m = Messung; b = Beobachtung; O = Facette in FDW bzw. PCK explizit eingebunden; – = Facette in FDW bzw. PCK nicht eingebunden; (O) = Facette im Modell enthalten, aber nicht im Messinstrument

Dabei zeigt sich einerseits, dass sich viele dieser Studien auf das transformative Modell fachdidaktischen Wissens stützen und damit das fachdidaktische Wissen als eigenständige Wissensdimension in Abgrenzung zum Fachwissen und pädagogischen Wissen auffassen (Abb. 6). Damit fehlt dieser Wissensdimension aber der von Hashweh (2005) als besonders wichtig erachtete Themenbezug, was gemäss van Dijk & Kattmann (2010) problematisch ist. Gess-Newsome (2015, S. 40) hat sich deshalb mit der durch das Konsensmodell erfolgten weiteren Klärung neuerdings von der Unterteilung in transformative und integrative Modelle gelöst (Kapitel I, 2.5.2). Sie unterscheidet neu zwischen dem themenspezifischen fachdidaktischen Professionswissen (TSPK) und dem themen- und kontextspezifischen PCK, zwei Formen fachdidaktischen Wissens und erachtet diese insgesamt als eine eigenständige Wissensdimension. Andererseits wird aufgrund der ausgewählten Wissensfacetten deutlich, dass zwischen den Studien ein Konsens besteht hinsichtlich der beiden zentralen Facetten fachdidaktischen Wissens von Shulman (1986), nämlich Instruktionsstrategien und Repräsentationen sowie Schülerverständnis (cf. van Driel, Verloop & de Vos, 1998). Diese beiden Wissensfacetten werden in der naturwissenschaftsdidaktischen Forschung als besonders relevant für das FDW bzw. für PCK eingestuft (Blömeke, Kaiser, et al., 2008; Grossman, 1990; Tamir, 1988; Tepner et al., 2012). Dies insbesondere, weil sie die Lehrenden befähigen, „die Probleme der SuS [Schülerinnen und Schüler] zu antizipieren und darauf angemessen zu reagieren“ (van Dijk & Kattmann, 2010, S. 8). Die scheinbare Einigkeit in Bezug auf die beiden Kernfacetten Shulmans (1986), wie auch bezüglich gewisser anderer Facetten sollte aufgrund der dahinterliegenden, unterschiedlichen Interpretationen der Facetten jedoch mit Vorsicht gedeutet werden (Gramzow et al., 2013). Die Kernfacette Schülerverständnis wird gemäss Kirschner et al. (2016, S. 1347) auf der einen Seite als Wissen zu alternativen Wissenskonzepten von Schülerinnen und Schüler interpretiert, auf der anderen Seite aber weitreichender verstanden als das Wissen über die Lernenden und das Lernen schlechthin. Hinter den gleichen Facettenbezeichnungen verschiedener Autoren stecken demzufolge durchaus unterschiedliche Konstrukte, die vordergründig nicht sichtbar sind.

Die Übersichtsdarstellung zeigt ausserdem, dass die Herangehensweisen bei der Erfassung des fachdidaktischen Wissens (FDW bzw. PCK) als Wissensdimension je nach Zielsetzungen der Studien verschieden sind (Tab. 3). Studien, die das fachdidaktische Wissen im Sinne von *PCK on action* und in gewisser Weise als Disposition auffassen, fokussieren fast ausschliesslich nur die beiden Kernfacetten Shulmans (1986) und messen sie mehrheitlich mit Hilfe von Paper-Pencil-Tests (e.g. Grossschedl, Neubrand, et al., 2015; Park & Suh, 2015) oder auch mit Hilfe von Vignetten (Brovelli, Bölsterli, Rehm & Wilhelm, 2013; Riese, 2009). Studien, die PCK handlungsnaher als *PCK in action* interpretieren,

beschreiben diese Wissensdimension breiter und erfassen sie mit Hilfe von qualitativen und teilweise auch quantitativen Beobachtungen wie Protokollen, Interviews oder auch Unterrichtsbeobachtungen (e.g. Gess-Newsome et al., 2017). Die hier beschriebenen, unterschiedlichen Erhebungsverfahren widerspiegeln sich auch in den Stichprobengrößen, die bei den Tests aus testökonomischen Gründen mehrheitlich grösser sind (cf. Kirschner et al., 2017).

Die vorliegende Untersuchung fokussiert das FDW, das in Unterrichtsplanungen beobachtet werden kann. Das FDW wird in Anlehnung an Modellierungen von PCK in Bezug auf die Wissensfacetten eindimensional konzeptualisiert. Durch den inhaltlich vorgegebenen Fokus der Erkenntnisgewinnung fällt die Teildimension der Inhaltsbereiche weg (cf. Riese, 2009). Auch werden Überlegungen zu den Wissenstypen getätigt, diese aber wie in der COACTIV-Studie nicht im Modell berücksichtigt. Die Modellierung der Wissensfacetten erfolgt auf der Grundlage des für die Naturwissenschaftsdidaktik wichtigen Planungsmodells der «Didaktischen Rekonstruktion» (Kattmann, Duit, Gropengiesser & Komorek, 1997; Kap. I, 2.8.4). Entsprechend werden neben den Kernfacetten Shulmans (1986) auch weitere, für den Planungsprozess wichtige Facetten einbezogen und operationalisiert (Kap. I, 3.2.4). Mit diesem Strukturmodell soll das fachdidaktische Planungswissen zum Experimentieren im Sinne der Erkenntnisgewinnung erfasst werden. Welche Wissenstypen fachdidaktischen Planungswissens sich mit dieser Untersuchung erheben lassen, soll im nächsten Abschnitt diskutiert werden.

VI. Wissenstypen zum FDW und PCK

Für das fachdidaktische Wissen (FDW bzw. PCK) werden im Spannungsfeld von Wissen und Können je nach Forschungs- und Theorieperspektive verschiedene Wissenstypen unterschieden. Einige Konzeptionen stützen sich dabei auf kognitionspsychologische Theorien, andere wiederum basieren auf einer epistemologischen Sichtweise, die das Wissen in seiner Funktion bei der Ausübung einer Aufgabe charakterisieren (de Jong & Ferguson-Hessler, 1996, S. 106). Insgesamt sind die Konzeptionen zu den Wissenstypen entsprechend vielfältig.

Zwei Aspekte scheinen jedoch allen Konzeptionen gemeinsam zu sein:

- Das fachdidaktische Wissen wird implizit oft als Voraussetzung für das Ausüben der Lehrtätigkeit im Sinne einer Handlungsressource angesehen (Vogelsang & Reinhold, 2013a).
- Das fachdidaktische Wissen wird selten nur als ein Wissenstyp aufgefasst (Stender, 2014).

In Bezug auf die Konzeptionalisierungen fachdidaktischer Wissenstypen sind in der Forschungsliteratur häufig zweidimensionale Modelle vorzufinden, die u.a. zwischen Wis-

senstypen wie formal-theoretischem und praktischem Wissen und Können (e.g. Baumert & Kunter, 2011a, S. 35; Fenstermacher, 1994) als auch zwischen deklarativem und prozeduralem Wissen unterscheiden (e.g. Baumert & Kunter, 2006; Schmelzing, Wüsten, Sandmann & Neuhaus, 2010). Die Unterscheidung zwischen deklarativem und prozeduralem fachdidaktischem Wissen ist insbesondere in naturwissenschaftsdidaktischen Forschungsprojekten weit verbreitet. In gewissen Untersuchungen wird dieses duale Modell noch mit weiteren Wissenskomponenten wie z.B. dem reflexiven Wissen (Schmelzing, 2010), dem konditionalen Wissen (e.g. Jüttner & Neuhaus, 2012; Kirschner, Borowski, Fischer, Gess-Newsome & von Aufschnaiter, 2016; Tepner et al., 2012) oder dem strategischen und schematischen Wissen (Shavelson, Ruiz-Primo & Wiley, 2005) ergänzt. Dabei resultieren drei- bis vierdimensionale Konstrukte, entsprechend vielfältig sind die Modellierungen. Eine Vielzahl an Modellierungen und Interpretationen sind aber bei genauerer Betrachtung auch schon beim dualen Modell selbst zu verzeichnen. Zur besseren Einordnung der vorliegenden Studie werden diese beiden Wissenstypen, d.h. das deklarative und prozedurale Wissen, zuerst aus kognitionswissenschaftlicher und epistemologischer Sicht beleuchtet, um vor diesem Hintergrund auf deren Verwendung in der naturwissenschaftsdidaktischen Forschung und der vorliegenden Studie einzugehen.

In der *kognitiven Neurowissenschaft* wird zwischen deklarativem und nondeklarativem Gedächtnis unterschieden, wobei letzterem auch das prozedurale Gedächtnis zugeordnet wird (Jäncke, 2017). Gemäss Jäncke kann diese Unterscheidung „als der Unterschied zwischen Wissen und Können“ bezeichnet werden (ebd., S. 443). Sie wird durch die Theorie des deklarativen Gedächtnisses gestützt. Diese schreibt dem mesiotemporalen Gedächtnissystem eine wichtige Funktion in der bewussten Verarbeitung u.a. von räumlichen und verbalen Informationen zu, währenddem dieses Gedächtnissystem am nondeklarativen, impliziten Gedächtnis nicht beteiligt ist (Squire, Stark & Clark, 2004). Die Wiedergabe von Gedächtnisinhalten im nondeklarativen Gedächtnis erfolgt demnach unbewusst, automatisch und ohne Willensanstrengung (Jäncke, 2017). Das bedeutet, dass die hier beschriebene Unterscheidung beider Gedächtnis- und Wissenstypen auf dem Kriterium des Bewusstseins basiert, das sich in explizitem und implizitem Wissen äussert. Neuere Nachweise zeigen allerdings, dass auch unbewusste Informationen über den dem mesiotemporalen System angehörenden Hippocampus gelernt und abgerufen werden (Henke, 2010). Damit wird das weit akzeptierte Unterscheidungskriterium des Bewusstseins etwas in Frage gestellt. Jäncke (2017, S. 487) vermutet, dass der „Hippocampus für die Verknüpfung und Assoziation von Informationen während der Konsolidierung verantwortlich [ist]“, losgelöst davon, ob es sich um bewusste oder unbewusste Lernprozesse handelt.

Anderson (2013) stützt sich aus *kognitionspsychologischer Perspektive* ebenfalls auf die von Squire, Stark und Clark (2004) vorgenommene Unterscheidung der Gedächtnistypen. Das deklarative Wissen wird hier als explizites Wissen, als ein Wissen über verschiedene Sachverhalte verstanden, während das prozedurale Wissen ebenfalls als implizites Wissen beschrieben wird, das sich darauf bezieht, wie man bestimmte Aufgaben ausführt. Bei der Entwicklung von Expertise ändert sich gemäss Anderson das Ausmass, in dem man sich auf deklaratives versus prozedurales Wissen verlässt, drastisch (Anderson, 2013, S. 199). Mit dem Prozess der Prozeduralisierung wird die explizite Verwendung deklarativen Wissens von der unmittelbaren Anwendung prozeduralen Wissens abgelöst (Anderson, 2013, S. 201).

Ansätze der *Informationsverarbeitungstheorie* fassen gemäss Oberauer (1993) das deklarative wie auch das prozedurale Wissen als zwei Kategorien der Kognition, nicht der Handlung auf. Mit deklarativem Wissen ist die Speicherung von Fakten im Gedächtnis gemeint, während das prozedurale Wissen als kognitiver Umgang und Manipulation mit Fakten beschrieben und damit von der konkreten Handlung abgegrenzt wird.

Hinsichtlich der Unterscheidung deklarativen und prozeduralen Wissens wird häufig auch Bezug auf den Philosophen Ryle (2009, 1949) genommen. Dieser unterscheidet zwischen Wissen im Sinne von *knowing that*, d.h. dem Wissen, das verbal und symbolisch ausgedrückt werden kann und Wissen im Sinne von *knowing how*. Letzteres beschreibt er wie folgt:

„Knowing how, then, is a disposition, but not a single-track disposition like a reflex or a habit. Its exercises are observances of rules or canons or the applications of criteria, but they are not tandem operations of theoretically avowing maxims and then putting them into practice.” (Ryle, 2009, S. 34)

Ryle fasst den Wissenstyp *knowing how* entsprechend der Informationsverarbeitungstheorie als Disposition auf. Gemäss Oberauer (1993) kritisiert Ryle aber die in dieser Theorie vorgenommene Trennung zwischen Geist und Körper und in diesem Zusammenhang auch die postulierte kausale Wechselwirkung, die zwischen diesen vorherrscht. Erklärungen bestimmter Verhaltensweisen sind aus seiner Sicht nur dispositional zu verstehen, d.h. das Können verweist auf Dispositionen, die den Menschen zugeschrieben werden (Oberauer, 1993, S. 32).

Neuweg (2014) fokussiert für die Konzeptionalisierung des „Lehrerwissens“ die gesamte Breite der Lehrerbildung und modelliert die Wissensformen eher aus einer *epistemologischen Sichtweise*. Dabei unterscheidet er innerhalb des Professionswissens von Lehrpersonen insgesamt drei Wissenstypen (Abb. 9)

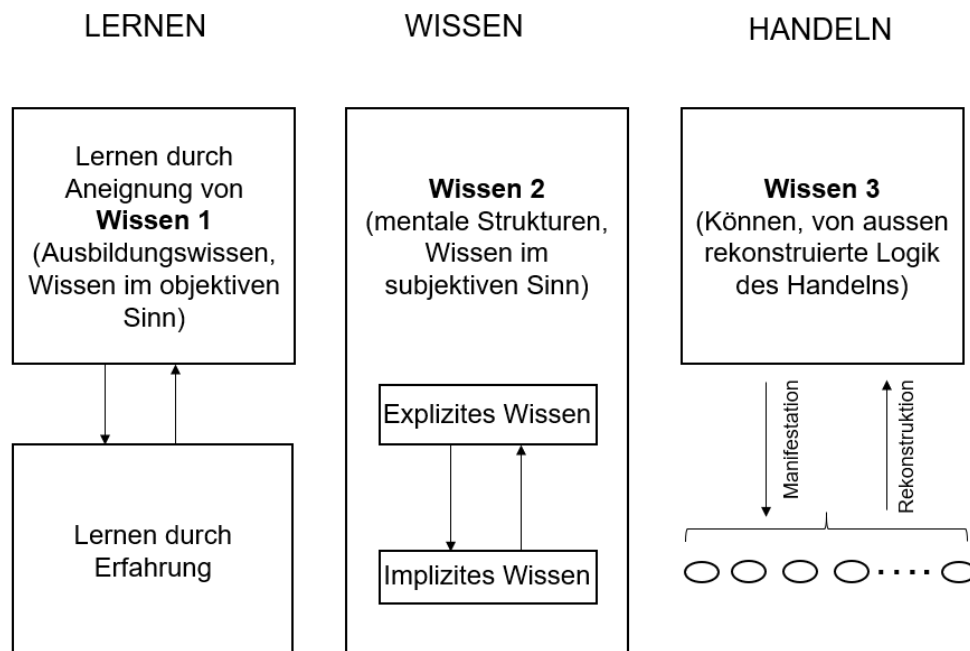


Abbildung 9: Konzepte des „Lehrerwissens“

Mit Wissen 1 bezeichnet er das objektive Wissen, mit Wissen 2 das subjektive Wissen und mit Wissen 3 das Können im Sinne der Performanz. Während Wissen 1 gemäss diesem Modell in der Ausbildung erworben werden kann, entspricht das Wissen 2 einem psychologischen Konstrukt im Sinne einer Erzeugungsgrundlage für kompetentes Verhalten (z.B. Schema, Skript, deklaratives und prozedurales Wissen) und Wissen 3 dem Können, dass sich im beobachtbaren Handeln zeigt. Neuweg nimmt in seinem Modell des „Lehrerwissens“ nur indirekt Bezug zum deklarativen bzw. prozeduralen Wissen und definiert die Wissenstypen nach deren Erwerb und Anwendung in der Lehrpersonenausbildung bzw. auch nach deren Explizierbarkeit (Neuweg, 2014). Wissen 1 erachtet er über die direkte Erfragung von objektivem Wissen in Form gängiger Test als zweifellos möglich. Wissen 2 lässt sich gemäss seinen Ausführungen durch Befragung oder über situierte Aufgaben testen. Rekonstruierbar ist es auch durch die Analyse von Denkprozessen v.a. im Rahmen von Unterrichtsplanungen, aber auch über *stimulated recalls* (cf. Kap. I, 2.4.2). Das Wissen 3 zeigt sich durch Beobachtungen im Unterricht oder an sehr unterrichtsnahen Aufgaben, wobei dies nur über das Wissen des Forschers erfolgen kann, der die Logik des Handelns von aussen betrachtet (Neuweg, 2014, S. 585).

In der naturwissenschaftsdidaktischen Forschungsliteratur gibt es sowohl für das fachdidaktische Wissen (FDW) wie auch für *pedagogical content knowledge* (PCK) Unterschiede in den Konzeptionen dieser Wissenstypen. Diese sollen in der Folge vorgestellt werden.

VII. Wissenstypen zum FDW bzw. PCK – Sicht der Naturwissenschaftsdidaktik

In Bezug auf das FDW betreffen die Unterschiede die Kongruenz der verschiedenen Konzeptionalisierungen. Die Konstrukte deklarativen Wissens sind dabei weitgehend ähnlich und decken sich mit den Auffassungen der kognitiven Neurowissenschaft, Kognitionspsychologie und den weiter beschriebenen Konzepten. Das deklarative fachdidaktische Wissen wird auch hier im Sinne von *knowing that* verstanden als explizites, d.h. verbalisierbares Faktenwissen über Begriffe, Regeln und Prinzipien, wie auch als konzeptionelles Wissen (e.g. Gramzow, Riese & Reinhold, 2013; de Jong & Ferguson-Hessler, 1996; Kirschner, Borowski, Fischer, Gess-Newsome & von Aufschnaiter, 2016; Schmelzing et al., 2013; Shavelson, Ruiz-Primo & Wiley, 2005). Beim prozeduralen Wissen hingegen bestehen grössere Unterschiede in den Konzeptionalisierungen und den jeweils zugeschriebenen Bedeutungen dieses Wissenstyps. Obgleich es bei allen Konzeptionalisierungen um ein Wissen wie (*knowing how*) geht, sind Unterschiede in der Handlungsnähe und der Explizierbarkeit dieses Wissens festzustellen. Zum einen wird unter prozeduralem fachdidaktischem Wissen der kognitive Umgang mit deklarativem Wissen verstanden (Grossschedl, Neubrand et al., 2015; Jüttner, Boone, Soonhye & Neuhaus, 2013; Kirschner et al., 2017). Dabei geht es im Sinne von Oberauer (1993) um Kenntnisse zu Handlungen und Prozessen und deren Umsetzung, d.h. um ein Wissen, wie man in bestimmten Situationen agieren soll. In gewisser Weise wird hier das prozedurale Wissen als Disposition und Grundlage für das Handeln beschrieben. Neuweg spricht hier auch von handlungsvorbereitendem Wissen (Neuweg, 2005, S. 20). Zum anderen wird das prozedurale fachdidaktische Wissen als Wahrnehmungs-, Verarbeitungs- und Handlungsrouninen aufgefasst, d.h. als Fähigkeiten, Fertigkeiten oder *skills*, die sich in situationsspezifischen Anwendungen entwickeln und zeigen und gegebenenfalls auch nicht vollständig explizierbar sind (e.g. Schmelzing, Wüsten, Sandmann & Neuhaus, 2010; Tamir, 1988). Hier einzuordnen wäre auch die Beschreibung prozeduralen Wissens von Stender (2014), die diesen Wissenstyp als Ergebnis einer Entwicklung aufgrund einer aktiven Auseinandersetzung mit einzelnen Unterrichtssituationen definiert (Stender, 2014, S. 21).

Mayer, Grube und Möller (2008, S. 65) verstehen unter prozeduralem Wissen gar die Prozesse des wissenschaftlichen Denkens bzw. die Prozeduren naturwissenschaftlicher Erkenntnisgewinnung und bezeichnen diese Wissensressource auch als Methodenwissen.

Dieses handlungsnähere Verständnis prozeduralen Wissens entspricht eher dem Ansatz der kognitiven Neurowissenschaft, wobei dieses Wissen in der naturwissenschaftsdidaktischen Forschung als explizierbar angenommen wird.

Für das etwas enger gefasste *pedagogical content knowledge* (PCK) finden sich dagegen in der Literatur kaum Modellierungen verschiedener Wissenstypen. Gemäss Alonzo und Kim (2016) liegt der Fokus vieler Untersuchungen auf dem deklarativen PCK, ohne dass dies explizit erwähnt wird. Mit der deklarativen Wissensdimension werden aber die verschiedenen Herausforderungen der Praxis nicht abgebildet. Park und Oliver verdeutlichen dies mit der Aussage „PCK encompasses both teachers’ understanding and enactment“ (Park & Oliver, 2008, S. 263). Alonzo und Kim (2016, S. 1263) unterscheiden deshalb in ihrer Studie zwischen dem deklarativen PCK und einem dynamischeren PCK, das eine stärkere Anbindung zur Praxis beinhaltet. Ähnlich, aber nicht genau gleich, wird PCK im Konsensmodell von Gess-Newsome (2015) zum einen als kognitive, kontextspezifische Disposition im Sinne von Wissen oder *knowledge base* (PCK) beschrieben, zum anderen als PCK, das sich in der direkten Handlung zeigt, also als *PCK and skill* (PCK&S) (cf. Kap. I, 2.5.2).

Die bisherigen Ausführungen zu den Konzeptionalisierungen deklarativen und prozeduralen Wissens zeigen eine Vielzahl an Konstrukten und Begriffsdefinitionen, die auf unterschiedliche Forschungstraditionen zurückgehen. Insbesondere Konstrukte zum prozeduralen Wissen sind z.T. schwer miteinander vereinbar. Während das prozedurale Wissen in der kognitiven Neurowissenschaft und der Kognitionspsychologie strenggenommen als grösstenteils implizites Wissen im Sinne von Fähigkeiten und Fertigkeiten sowie automatisierten Handlungsrouinen verstanden wird, wird mit dem Begriff des prozeduralen Wissens in der fachdidaktischen Forschung eine situationsbezogene Handlungsnähe ausgedrückt, die je nach Konzeptionalisierung eher deklarativen und expliziten bzw. prozeduralen und expliziten oder impliziten Charakter aufweist. In den Kognitionswissenschaften wird das Wissen nach seiner Rolle in der Wissensverarbeitung beschrieben, in der naturwissenschaftsdidaktischen Forschung eher nach der inhaltlichen Ausrichtung des Wissens im gesamten Prozess jeglicher unterrichtlichen Tätigkeit. Letzteres findet sich auch im Konsensmodell von Gess-Newsome (2015) wie auch in der Modellierung des „Lehrerwissens“ von Neuweg (2014).

In beiden Modellen wird das Wissen im Sinne unterscheidbarer Facetten des Denkens und Handelns (Neuweg, 2014) in Bezug auf den gesamten Prozess der Unterrichtstätigkeit modelliert. *PCK on action* im Konsensmodell entspricht dabei dem persönlichen, situations- und kontextspezifischen fachdidaktischen Wissen, was mit dem Wissen 2 von Neuweg (2014) vergleichbar ist. Dieses Wissen ist handlungsvorbereitend. Neuweg (2014) weist ihm explizite wie implizite Anteile zu. Demgegenüber beschreibt *PCK in action* das Können, was dem Wissen 3 der Modellierung Neuwegs (2014) entspricht.

Die vorliegende Untersuchung fokussiert das fachdidaktische Wissen zur Planung von Unterricht zum experimentellen Handeln im Sinne der Erkenntnisgewinnung. Gemäss Stender, Geller, Neumann und Fischer (2013) kann vermutet werden, dass explizites Wissen im Unterrichtsplanungsprozess wirksam wird. Es handelt sich dabei um persönliches, situations- und kontextspezifisches fachdidaktisches Wissen, das expliziert wird (Gess-Newsome, 2015; Neuweg, 2014). In Anlehnung an Oberauer (1993) wird dieses in Unterrichtsplanungen beobachtbare Planungswissen in dieser Untersuchung aber als Kategorie der Kognition aufgefasst. Es wird als handlungsvorbereitendes Wissen im Sinne eines im Rahmen des Planungsprozesses erfolgenden kognitiven Umgangs und einer Manipulation mit Fakten angesehen. Das Können, d.h. die Fähigkeit zur integrierten Anwendung fachdidaktischen Wissens, lässt sich bis zu einem gewissen Grad aus der Qualität der Unterrichtsplanung ableiten.

Die durch den Paradigmenwechsel in der Professionsforschung der Lehrpersonen intensive Beschäftigung mit dem Professionswissen (Kapitel I, 2.3) impliziert einen grundsätzlichen Zusammenhang zwischen Wissen und Handeln (Cauet, 2016, S. 21). Könnerschaft ist aber gemäss Neuweg (2005) „niemals blossse Wissensapplikation“ [, sondern] „schliesst die Kunst der Kontextualisierung dieses Wissens auf besondere Fälle ein“. Nach wie vor nicht ausreichend geklärt ist, ob und wie das explizierbare Professionswissen der Lehrpersonen beim Planen von Unterricht und beim Unterrichten wirksam wird. Denkbar ist, dass das Fachwissen bzw. das fachdidaktische Wissen zwar eine notwendige, aber nicht hinreichende Disposition für ein erfolgreiches Handeln im Unterricht darstellt (e.g. Abell, 2007; Fischler, 2008; Lederman & Gess-Newsome, 1992; Oser & Oelkers, 2001). Das Können einer Lehrperson, d.h. ihre professionelle Kompetenz, sollte sich aber zumindest in der Qualität des von ihr (geplanten oder) realisierten Unterrichts, d.h. in der Performanz zeigen (e.g. Helmke, 2012; Neuhaus, 2007; Vogelsang & Reinhold, 2013). Dies bedingt, dass Kriterien für einen erfolgreichen Unterricht formuliert werden müssen und die Qualität des Unterrichts betrachtet werden sollte. Das nächste Kapitel widmet sich vor diesem Hintergrund dem Thema der Unterrichtsqualität.

2.6 Unterrichtsqualität – Einfluss des Professionswissens von Lehrpersonen

Unterrichtsqualität wird von Einsiedler (2002, S. 195) definiert als „Bündel von Unterrichtsmerkmalen, die sich als Bedingungsseite (oder Prozessqualität) auf Unterrichts- und Erziehungsziele („Kriterienseite“ oder Produktqualität) positiv auswirken, wobei die Kriterienseite überwiegend von normativen Feststellungen bestimmt und der Zusammenhang

von Unterrichtsmerkmalen und Zielerreichung von empirischen Aussagen geleitet ist“. Der Begriff Unterrichtsqualität entspricht damit einer normativen Anschauung bzw. Wertung guten Unterrichts, die über eine reine Beschreibung von Eigenschaften („Qualitäten“) des Unterrichts hinausgeht (Clausen, 2002, S. 15). Vor diesem Hintergrund kann Unterrichtsqualität zum einen mit Hilfe normativer Vorstellungen zum guten Unterricht, u.a. auch in Form von Lehrplänen und Bildungsstandards ermittelt werden. Zum anderen lässt sich Unterrichtsqualität bestimmen anhand von Wirkungen (oder Effekten) des erteilten Unterrichts, die über Leistungen der Schülerinnen und Schüler oder auch mittels deren Einstellungen, Haltungen und Werten erfasst werden (Ditton, 2009). Hierzu lassen sich zwei Forschungsstränge zuordnen, einerseits die Unterrichtsqualitätsforschung im Sinne des «Prozess-Mediations-Produkt-Paradigmas», andererseits die Lehrerprofessionalitätsforschung bzw. das «Expertenparadigma» (cf. Kapitel I, 2.3.2 & I, 2.3.3). Beide Forschungsstränge haben sich heute in der Lehr- und Lernforschung durchgesetzt, da sich mit diesen beiden Untersuchungsfokussen die potenzielle Wirkkette von der Lehrperson über den Unterricht zu den Schülerinnen und Schülern in ihrer Gesamtheit untersuchen lässt (vergleichende Übersicht siehe Kotzebue & Neuhaus, 2016). Berliner (2005) vertritt in diesem Zusammenhang die Ansicht, dass zur Bestimmung der Unterrichtsqualität die normative Perspektive zwingend miteinbezogen werden sollte und bezeichnet die Summe eines guten und effektiven Unterrichts als qualitätsvollen Unterricht (Berliner, 2005, S. 207). Für die Ermittlung bedeutsamer Qualitätsmerkmale sind in der Forschungsliteratur verschiedene Herangehensweisen gewählt worden. Der folgende Abschnitt schafft dazu eine Übersicht.

2.6.1 Unterrichtsqualität – die Suche nach Qualitätsmerkmalen

Damit Aussagen über die Qualität von Unterricht gemacht werden können, sollte der schulische Unterricht, ein besonderes Tätigkeitsfeld, in seiner Gesamtheit analysiert werden (Reusser & Pauli, 2010). Verschiedene Herangehensweisen sind hierzu in der Unterrichtsqualitätsforschung gewählt worden. Zum einen sind Listen der wichtigsten Qualitätskriterien auf der Basis von Literaturreviews erstellt worden (Brophy, 2000; Helmke, 2012). Zum anderen sind Effekte einzelner Qualitätsmerkmale ermittelt worden (e.g. Jatzwauk, Rumann & Sandmann, 2008; Wadouh, Sandmann & Neuhaus, 2009) oder aber es sind systematische Ordnungen der Unterrichtsqualitätsmerkmale in Form von Reviews (Dorffner, Förtsch & Neuhaus, 2017) oder in Metaanalysen zusammengetragen worden (Hattie, 2009; Kyriakides, Christoforou & Charalambous, 2013; Seidel & Shavelson, 2007). Ausserdem sind auch Rahmenmodelle entwickelt worden, die sowohl die Einflüsse der Lehrperson und der individuellen Voraussetzungen der Schülerinnen und Schüler als auch der Unterrichtsgestaltung im Kontext der Gesellschaft und Schule auf die Unterrichtserträge

berücksichtigen. Dazu gehören im deutschsprachigen Raum die Angebots-Nutzungs-Modelle unterrichtlicher Wirkungen von Fend (e.g. 2002, 2008) und Helmke (e.g. 2012). Gemäss Seidel (2014b) lassen sich in diesen Modellen drei Ebenen unterscheiden. Das sind die Angebotsstrukturen im Sinne von Lerngelegenheiten, die Nutzungsformen dieser Lerngelegenheiten sowie die Lernergebnisse. Diese sogenannten «Prozess-Mediations-Produkt-Modelle» berücksichtigen die verschiedenen Einflussfaktoren auf Angebots- und Nutzerseite, ebenso wie die ko-konstruktiven Prozesse und Aktivitäten der Lernenden (Reusser & Pauli, 2010; cf. Kapitel I, 2.3.2). „Unterricht wird in diesem Rahmen als mehr oder weniger qualitätsvolles Angebot von Lerngelegenheiten verstanden, das von den Lernenden auf mehr oder weniger qualitätsvolle Weise genutzt wird“ (Reusser, 2009, S. 222). Aus der Perspektive der Forschung interessieren auf der Seite des Angebots „die Bestimmung von Kompetenzen bei Lehrenden und die Identifikation von Unterrichtsmerkmalen“ im Sinne des «Expertenparadigmas» (Seidel, 2014b, S. 860; cf. Kapitel I, 2.3.3). Aus Sicht der Nutzung liegt der Fokus auf „Theorien zu Prozessen des Lernens und deren Relation zu Unterrichtsmerkmalen“ gemäss «Prozess-Mediations-Produkt-Paradigma» (Seidel, 2014b, S. 854; cf. Kapitel I, 2.3.2). Dabei wird der Zusammenhang zwischen Unterricht und dessen Output auf der Ebene der Schülerinnen und Schülern betrachtet und es werden Unterrichtsmerkmale eruiert, die hinsichtlich bestimmter messbarer Kriterien für die Schülerinnen und Schüler gewinnbringend sein könnten (Dorfner et al., 2017). Die Effekte von Unterricht werden aber nicht als Wirkung, sondern aufgrund der Interdependenzen im Sinne einer Folgeabschätzung abgebildet (Klieme, 2006). Diese Angebots-Nutzungs-Modelle sind in Bezug auf unterschiedliche Unterrichtskontexte und -fächer weiter ausgeschärft und verändert worden. Ein Modell, das zusätzlich auch die systemische Bedingtheit miteinbezieht, ist von Reusser und Pauli (2010) entwickelt worden. Welche (fach-)didaktischen Qualitätsmerkmale bisher auf diesen Grundlagen ermittelt und zusammengetragen worden sind, ist Gegenstand des nächstfolgenden Abschnitts.

2.6.2 Unterrichtsqualität – (fach-)didaktische Qualitätsmerkmale

Die Unterrichtsforschung hat viele Jahre vor allem nach empirisch belegbaren allgemeinen, fächerübergreifenden Merkmalen guten Unterrichts gesucht (Baer et al., 2011). Hierbei sind relevante Merkmale erfolgreichen Unterrichts identifiziert worden, die einen positiven Einfluss auf die Schülerinnen und Schüler haben sollen.

Gemäss Ditton (2009) lassen sich diese Merkmale drei Bereichen oder Dimensionen zuordnen:

- die Unterrichts- und Klassenführung mit einer klar organisierten und gut strukturierten Instruktion,
- die Schülerorientierung und die damit verbundene Orientierung an den Lernpotenzialen und -bedürfnissen der Lernenden sowie
- die kognitive Aktivierung in Verbindung mit der Aufgabenkomplexität und der hohen Intensität fachlichen Lernens (Ditton, 2009, S. 180).

Diese drei fächerübergreifenden Qualitätsdimensionen stellen eine „sparsame, aber relativ umfassende Systematisierung der, im Unterricht zu beobachtenden, Lehr- und Lernprozesse dar“ (Kunter & Ewald, 2016, S. 16). Sie werden von anderen Autoren gestützt (Brophy, 2000; Dorfner et al., 2017; Hattie & Yates, 2015; Helmke, 2012; Klieme, 2006; Kunter & Trautwein, 2013; Reusser, 2009) und haben sich u.a auch etabliert, weil es gute Instrumente zu deren Erfassung gibt (e.g. Klieme & Rakoczy, 2008; Kunter & Baumert, 2006; Lipowsky et al., 2009). Fach- bzw. - inhaltspezifische Merkmale, die, so Helmke (2012, S. 169), „mit Sicherheit einen wesentlichen Teil der Unterrichtsqualität aus[machen]“, sind jedoch häufig vernachlässigt worden (Wüsten, Schmelzing & Sandmann, 2010). Dabei ist die Unterrichtsqualität und damit die Qualität unterrichtlichen Handelns stets aus der Sicht der Fachdidaktik wie auch aus Sicht der Allgemeindidaktik zu bestimmen (Reusser, 2009, S. 222). Neuhaus (2007) hat deshalb das vorangehend beschriebene Angebots-Nutzungs-Modell für den Biologieunterricht weiter ausdifferenziert. Das Modell unterscheidet zwischen fachunabhängigen und fachspezifischen Qualitätsmerkmalen auf Lehrer-, Schüler- und Unterrichtsebene (Abb. 10). Als fachspezifische Merkmale auf der Ebene der Lehrperson werden u.a. ein konzeptuelles, fachspezifisches Wissen sowie ein Handlungsrepertoire in Bezug auf fachspezifische Qualitätsmerkmale aufgeführt. Auf der Ebene des Unterrichts sind im Rahmenmodell zur Unterrichtsqualität u.a. die Berücksichtigung der Sachstruktur, der Umgang mit Schülerfehlern, die Fokussierung auf wesentliche Fachbegriffe sowie die Einbettung und Umsetzung von Experimenten erwähnt. Gemäss der Autorin kann davon ausgegangen werden, dass sich fachunabhängige und fachspezifische Merkmale nicht additiv ergänzen, sondern Interaktionseffekte bestehen. Es ist anzunehmen, dass fachspezifische Merkmale eine notwendige Bedingung für die Wirksamkeit fachunabhängiger Merkmale darstellen. Fachinhalte sind daher eine notwendige Voraussetzung für fachspezifische Unterrichtsqualitätsmerkmale (Dorfner et al., 2017, S. 267), die sich laut Neuhaus (2007, S. 249) „im Sinne der Output-Steuerung von Unterricht [...] am erzielten Ertrag identifizier[en lassen]“.

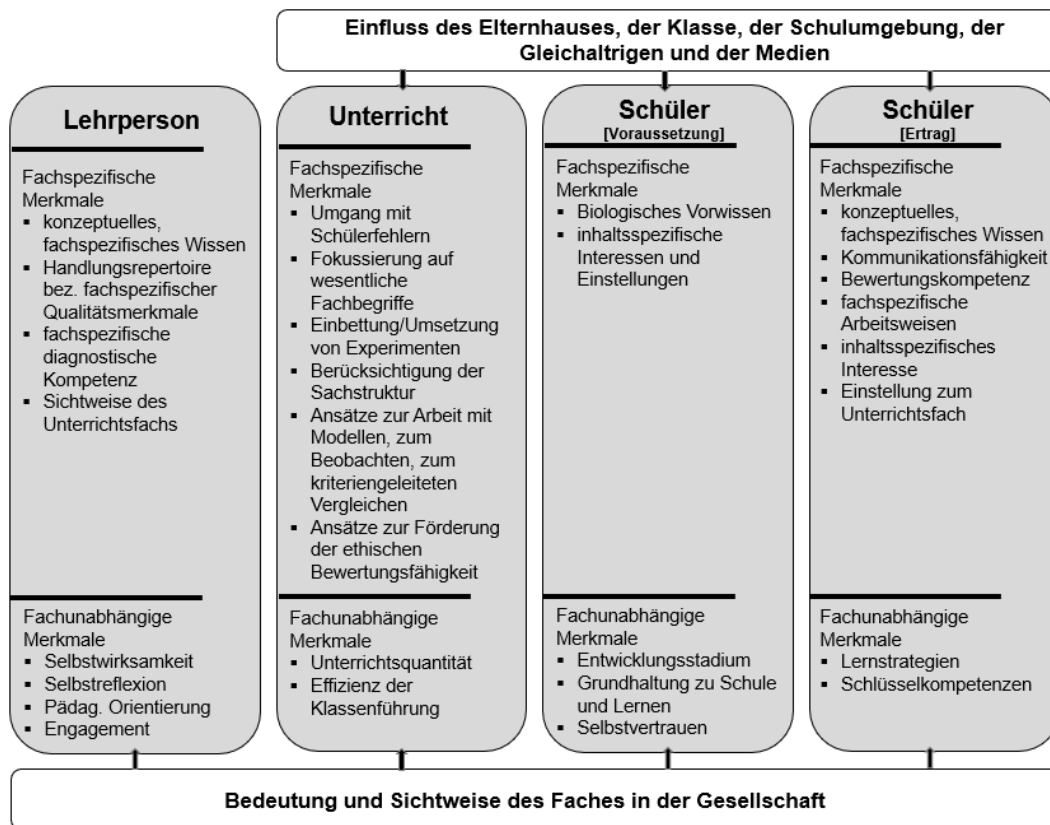


Abbildung 10: Fachspezifisches Rahmenmodell zur Unterrichtsqualität (Neuhaus, 2007)

In diesem Zusammenhang werden zwei Gruppen von Erträgen unterschieden. Zum einen Erträge, die zwar fachspezifisch erfasst werden, aber für alle oder die meisten Fächer gelten (Entwicklung einer konzeptionellen, fachspezifischen Wissensbasis, die Entwicklung inhaltsspezifischer Interessen etc.), zum anderen Erträge, die spezifisch für ein Fach bedeutungsvoll sind und für andere Fächer weniger. Letzteren kann gemäß Neuhaus (2007) u.a. auch die korrekte Anwendung spezifisch biologischer Denk- und Arbeitsweisen zugeordnet werden, wie das Beobachten, das Experimentieren, das kriterienorientierte Vergleichen sowie das Arbeiten mit Modellen (KMK, 2005). Diese Arbeits- und Denkweisen werden auch in anderen Arbeiten als Qualitätskriterien aufgeführt (Brovelli et al., 2013; Seidel et al., 2006; Seidel & Prenzel, 2004; Tepner et al., 2012; Wüsten, 2010).

Das Experiment und in diesem Zusammenhang das Experimentieren im Sinne der Einbettung und Einbeziehung experimentellen Denkens und Arbeitens wird als fachspezifisches Qualitätsmerkmal eingestuft, insbesondere auch im Biologieunterricht. Die Methode bzw. Organisationsform ist aber nicht per se eine Garantie für Unterrichtsqualität (Abrahams & Millar, 2008; Börlin, 2012; Hofstein & Lunetta, 2004). Wenn Unterrichtsqualität über Lernergebnisse definiert wird, dann müssen diese in Abhängigkeit von Lern- bzw. Unterrichtsprozessen beschrieben werden können (Fischer et al., 2003). Die Beschreibung der

Qualität von (Biologie-) Unterricht kann dabei grundsätzlich auf zwei Ebenen erfolgen. Zum einen auf der Ebene der Oberflächen- oder Sichtstruktur, die den eigentlichen Rahmen für die Unterrichtsprozesse bildet (Kunter & Trautwein, 2013), und übergeordnete Organisationsmerkmale wie Sozial- und Inszenierungsformen sowie Methoden und Medien miteinschließt (Reusser, 2009). Zum anderen auf der Ebene der Tiefenstruktur, welche die psychologisch-didaktischen Qualitätsdimensionen dieser Merkmale auf Prozessebene beschreibt und damit die Interaktionsebene zwischen Lehrenden und Lernenden (Kunter & Ewald, 2016; Kunter & Voss, 2011; Reusser, 2009). Letzterem können die vorangehend von Ditton (2009) erwähnten inhaltlichen Bereiche zugeordnet werden. Auch die Basismodelle von Oser & Baeriswyl (2001), die wichtige Lernprozesse unter der Perspektive unterschiedlicher Lernziele beschreiben, lassen sich in gewisser Weise hier zuteilen. Gemäss Hattie sowie Seidel & Shavelson wird den Merkmalen der Tiefenstruktur eine grössere Erklärungsmacht beigemessen, was die Lernerfolge der Lernenden betrifft (Hattie, 2009; Seidel & Shavelson, 2007). Sie werden oft als die eigentlichen Merkmale qualitätsvollen Unterrichts angesehen. Chomsky (1965) unterstreicht diesen Gedanken, indem er die Sichtstruktur als Ergebnis der Tiefenstruktur auffasst. Zur Bestimmung der Qualitätsdimensionen muss daher das Augenmerk auf die Prozessebene und damit die Ebene der Tiefenstruktur gelegt werden. Die Frage, welches Verständnis in der fachdidaktischen Forschungsliteratur dem Experiment zugrundegelegt wird und welche Qualitätsdimensionen im Rahmen des Experimentierprozesses von Bedeutung sind, wird im nächsten Kapitel näher erörtert.

2.7 Experimentieren – ein Qualitätsmerkmal naturwissenschaftlichen Unterrichts

Experimente nehmen im naturwissenschaftlichen Unterricht (Fischler, 2007; Schulz, 2011; Schulz, Riess, Wirtz & Barzel, 2012; Seidel et al., 2009; Seidel, 2014a; Tesch & Duit, 2004) insbesondere auch im Biologieunterricht, eine herausragende Stellung ein (Kizil & Kattmann, 2013). Die Bedeutung des Experimentierens zeigt sich auch in den normativen Vorgaben für den Biologieunterricht auf der Sekundarstufe I (Bögeholz, Joachim, Hasse & Hammann, 2016). Auf die Frage, was unter einem Experiment zu verstehen ist, haben die Wissenschaftsforschung und auch die naturwissenschaftsdidaktische Forschung keine eindeutigen, sondern mehrere Antworten (Höttecke & Riess, 2015). Auch erfüllt die Methode im Unterricht unterschiedliche Funktionen (Tesch, 2005) und ist, wie vorangehend erwähnt, nicht per se eine Garantie für qualitätsvollen Unterricht (Abrahams & Millar, 2008; Börlin, 2012; Hofstein & Lunetta, 2004). Mit der Einführung der neuen Bildungsstandards (EDK, 2011; KMK, 2004, 2005; NCR, 2012) haben sich die Anforderungen an das

Experimentieren zudem noch verändert und sind die naturwissenschaftlichen Denk- und Arbeitsweisen (*Scientific Inquiry*) ins Zentrum gerückt (Kapitel I, 2.1.2). Der nächste Abschnitt geht auf das experimentelle Handeln und *Scientific Inquiry* als Fachinhalt ein und beleuchtet diese Veränderungen.

2.7.1 Experimentelles Handeln – *Scientific Inquiry* als Fachinhalt

Der Einsatz von Experimenten ist früher stärker mit dem pädagogischen Ansatz *learning by doing* (Dewey, 1916) begründet worden. Das Experiment ist dabei hauptsächlich im Sinne eines Mediums zur Veranschaulichung eines Fachinhaltes oder zur Motivation eingesetzt worden. Heute steht beim Experimentieren, aber auch bei anderen naturwissenschaftlichen Denk- und Arbeitsweisen, die Entwicklung eines „grundlegenden Verständnisses von naturwissenschaftlichen Denk- und Arbeitsweisen“ (EDK, 2011; KMK, 2005; NCR, 2012) im Vordergrund (Kap. I, 2.1.1 & I, 2.1.2). Experimentieren im Sinne von *Scientific Inquiry* wird damit selbst zum Fachinhalt (Fazio, Melville & Bartley, 2010). Furtak, Seidel, Iverson & Briggs (2012) haben hierzu, ausgehend von Duschl (2008, 2003), vier kognitive Dimensionen definiert (*conceptual, epistemic, social, procedural*), mit denen sie das Konzept *Inquiry* klar von der weitverbreiteten Auffassung der reinen Hands-on-Aktivität abgrenzen. Im Kern geht es dabei um den Aufbau eines Verständnisses, wie das Wissen über die Welt mittels naturwissenschaftlicher Untersuchungen generiert werden kann (Gyllenpalm & Wickman, 2011a). Die Erforschung und Untersuchung naturwissenschaftlicher Phänomene (*Scientific Inquiry*) und die damit verbundene Auseinandersetzung mit dem Prozess der Erkenntnisgewinnung erlangen demzufolge eine grössere Bedeutung für den Unterricht (Lederman, Lederman & Antink, 2013; Millar, 2006). Dabei stehen in erster Linie prozessbezogene und fachmethodische Fähigkeiten und Fertigkeiten wie *doing science* und *knowing about science* (Hodson, 2014) im Fokus. Das Experiment, das häufig als gezielte „Frage an die Natur“ bezeichnet wird, erfährt somit eine neue Ausrichtung und spielt als methodisch kontrollierte Vorgehensweise eine wesentliche Rolle für die Auseinandersetzung mit Prozessen der Erkenntnisgewinnung (Kizil & Kattmann, 2013). Der Begriff des Experimentierens beinhaltet daher neu auch eine wissenschaftstheoretische und prozessbezogene Dimension, weshalb in der fachdidaktischen Literatur oft auch vom experimentellen Handeln im Unterricht die Rede ist. Börlin (2012) betont hierzu, dass dieser Terminus jeweils dann vorgezogen wird, wenn die Aktivität im Sinne naturwissenschaftlichen Problemlösens besonders hervorgehoben werden soll. Im englischsprachigen Raum wird auch häufig der Begriff *practical work* verwendet. Aufgrund des neuen Fokus wird das experimentelle Handeln selbst zum Fachinhalt (Börlin, 2012; Capps & Crawford, 2013a; Hasse, Joachim, Bögeholz & Hammann, 2014). Mit dieser neuen

Ausrichtung auf das experimentelle Handeln als Fachinhalt im Sinne von *Scientific Inquiry* stellt sich die Frage, wie experimentelle Kompetenzen in diesem Zusammenhang tatsächlich beschrieben und modelliert werden können. Der nächste Abschnitt nimmt dieses Thema auf.

2.7.2 Experimentelle Kompetenzen – verschiedene Auffassungen und entsprechende Modellierungsansätze zu *Scientific Inquiry*

Zur Beschreibung und Erfassung experimenteller Kompetenzen im Sinne von *Scientific Inquiry* im Unterricht liegt eine Vielzahl an Modellen und Messverfahren vor (Emden, 2011; Gut-Glanzmann, 2012; Rönnebeck, Bernholt & Ropohl, 2016; Schreiber, 2012). Im Zentrum dieser Modelle steht grundsätzlich die lerntheoretische Auffassung, dass die Prozesse der Erkenntnisgewinnung als komplexe Problemlöseprozesse zu verstehen sind. Diese lassen sich durch eine Abfolge spezifischer Prozeduren charakterisieren und werden von Situationsmerkmalen wie auch Personenvariablen beeinflusst (Klahr, 2000; Mayer, 2007; Roberts & Gott, 1999). Die verschiedenen Modellierungsansätze unterscheiden sich allerdings in Bezug auf das dem Experimentieren zugrundegelegte Verständnis. Während im naturwissenschaftlichen Unterricht dem Begriff Experiment aus Gründen der Generalisierung und sprachlichen Vereinfachung die unterschiedlichsten Tätigkeiten bis hin zur Erreichung technischer Ziele zugeordnet werden, wird das Experimentieren in der Biologie als ein kontrollierter Eingriff in einen zu erforschenden naturwissenschaftlichen Ablauf zur Generierung kausaler Zusammenhänge verstanden (Wellnitz, 2012, S. 45; Wellnitz & Mayer, 2013). Dieses Verständnis wirkt sich entsprechend auf die innere Abgrenzung der erforderlichen Teilkompetenzen aus, die sich ebenfalls unterscheiden (e.g. Gut, Hild, Metzger & Tardent, 2014). Im Zusammenhang mit den Messverfahren lassen sich deshalb zum Prozess naturwissenschaftlicher Erkenntnisgewinnung im Wesentlichen zwei Modellierungsansätze experimenteller Kompetenzen finden (Gott & Duggan, 2002; Roberts & Gott, 2003).

Ein erster Modellierungsansatz beschreibt das Experimentieren als integralen Problemlöseprozess. Objekte oder Zusammenhänge werden verglichen, klassifiziert, gemessen, untersucht oder konstruiert (Gut, Metzger, Hild & Tardent, 2014). Nach dieser Sichtweise unterscheiden sich verschiedene Experimentieraufgaben nach dem Typ des Problems, das durch das Experimentieren gelöst werden soll (Ruiz-Primo & Shavelson, 1996). Die Grundlage dieses Ansatzes bilden die *Concepts of Evicence*, d.h. ein Verständnis darüber, wie valide und reliable Daten gewonnen werden (Roberts & Gott, 2003, S. 116). Die Herangehensweisen bei der Problemlösung werden dabei in Anpassung an den Problemtyp modelliert und in der Regel mit Hilfe von Hands-on-Aufgaben erfasst. Aufgrund des im

Zentrum stehenden Verständnisses zum Erkenntnisprozess (*Concepts of Evidence*) lassen sich mehr Problemtypen analysieren (Roberts & Gott, 2003). Experimentelle Kompetenzen werden damit sehr viel weiter gefasst.

Im zweiten Modellierungsansatz wird der Experimentierprozess als Zusammenspiel unterscheidbarer Teilprozesse im Sinne eines idealisierten, hypothetisch-deduktiven Erkenntnisprozesses mit einer klaren Abfolge der Teilprozesse beschrieben (Gut, Metzger, et al., 2014). Die Güte der Problemlösung wird dabei durch die Qualität der Vorgehensweisen und das in diesem Zusammenhang vorhandene fachliche und strategische Wissen beeinflusst (Mayer, 2007). In der Literatur sind verschiedene Strukturierungen von Teilprozessen bzw. Experimentierphasen vorzufinden (cf. Emden, 2011, S. 13). Diese Teilprozesse lassen sich im Sinne eines idealtypischen, forschungslogischen, aber auch rekursiven Ablaufs in drei Hauptphasen zusammenfassen: (1) Fragestellungen und Hypothesenformulierung, (2) Planung und Durchführung der Untersuchung, (3) Datenanalyse und Reflexion der Untersuchung (Emden, 2011; Nowak, Nehring, Tiemann & Belzen, 2013; Wellnitz et al., 2017). Zahlreiche Modelle dieser Art sind domänenspezifisch und oft auch in Anlehnung an das *«scientific discovery as dual search»-model* (SDDS-Modell) von Klahr und Dunbar (1988) entwickelt worden, insbesondere auch Modelle zur Erfassung experimenteller Kompetenzen im Biologieunterricht (e.g. Hammann, Phan & Bayrhuber, 2008; Mayer et al., 2008; Nowak, Nehring, Tiemann & Belzen, 2013; Wellnitz & Mayer, 2013). Diese sind mit Hilfe von Paper-Pencil-Tests hinsichtlich der Teilprozesse und in diesem Zusammenhang möglicher, zugrundeliegender Dimensionen geprüft worden. Empirische Evidenzen für die Ausdifferenzierung prozessbezogener Teilkompetenzen konnten bislang aber nach wie vor nicht zufriedenstellend geklärt werden (Vorholzer, von Aufschnaiter & Kirschner, 2016; Wellnitz et al., 2017).

Die drei Hauptbereiche naturwissenschaftlicher Erkenntnisprozesse sind ungeachtet der vorangehend erwähnten empirischen Datenlage in internationalen und auch nationalen Standards verankert (EDK, 2011; KMK, 2005; Next Generation Science Standards [NGSS], 2013). In der Schweiz sind sie für die Sekundarstufe I im Handlungsaspekt „Fragen und Untersuchen“ der Grundkompetenzen (EDK, 2011) aufgeführt und in den Bildungsstandards in Deutschland im Kompetenzbereich Erkenntnisgewinnung (KMK, 2005) (Kapitel I, 2.4.3). Die Vermittlung von Kompetenzen zum Experimentieren im Sinne prozessbezogener Teilkompetenzen wird demzufolge als wichtiges Bildungsziel angesehen, insbesondere auch im Fach Biologie. Das dafür notwendige Professionswissen und die damit verbundenen professionellen Handlungskompetenzen sollen daher im Hinblick auf einen qualitätsvollen Biologieunterricht in der Lehrerbildung grundgelegt werden. Die

vorliegende Studie setzt hier an und baut auf dem Teilprozessansatz experimenteller Kompetenzen auf.

Die bisher beschriebenen Modellierungsansätze und entsprechenden Erhebungen sind ausgerichtet auf die Erfassung experimenteller Kompetenzen von Schülerinnen- und Schüler im Unterricht. Eine systematische Untersuchung der Lehrerkompetenzen ist im Vergleich dazu noch jung (Bögeholz et al., 2016). Entsprechend sind Modellierungsansätze zur Erfassung des fachdidaktischen Wissens zur Planung oder Umsetzung von Unterricht zum experimentellen Handeln von (angehenden) Lehrpersonen selten. Dies erstaunt, wenn man bedenkt, dass das, was Schülerinnen und Schüler lernen, stark von der Art und Weise, wie sie unterrichtet werden, abhängt (National Research Council [NRC], 1996, S. 28). Zwei neuere Studien, die sich mit der Analyse bzw. mit der Planung von Experimentalunterricht befassen und in diesem Zusammenhang entsprechende Erhebungsinstrumente entwickelt haben, seien hier vor dem Hintergrund der vorliegenden Studie kurz vorgestellt. Es handelt sich um die Projekte QuIP und ExMo.

Im Rahmen des trinationalen QuIP-Projekts (*quality of instruction in physics*) ist zur Erfassung der Qualität von videografiertem Experimentalunterricht im Fach Physik ein Erhebungsinstrument entwickelt worden (Börlin, 2012).

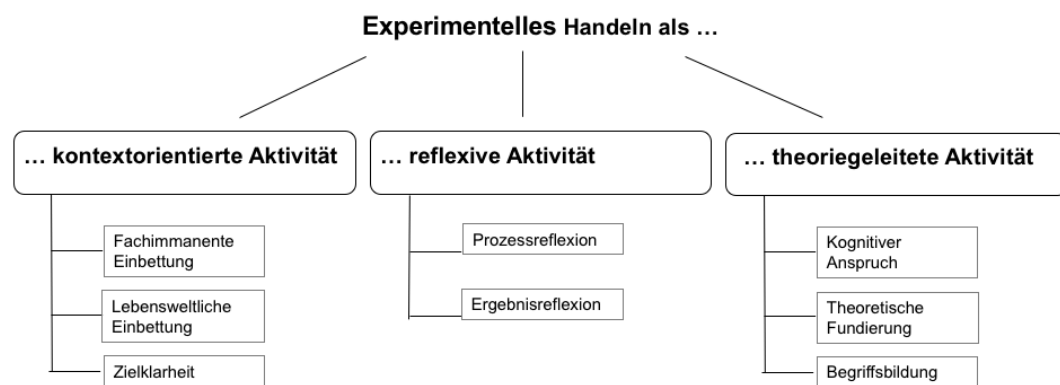


Abbildung 11: Struktur des Erhebungsinstruments aus dem QuIP-Projekt (Börlin, 2012, S. 57, von der Autorin angepasst)

Das dem Instrument zugrunde liegende Verständnis experimentellen Handelns ist eher weit gefasst. Das experimentelle Handeln wird als holistische Tätigkeit verstanden, welche die Vor- und Nachbereitung des Experimentierprozesses inkludiert (Börlin, 2012, cf. Millar, 2004; Tesch & Duit, 2004) und kontextorientiert, reflexiv und theoriegeleitet ist (Börlin, 2012, S. 35; Abb. 11).

Letzteres bildet die Grundlage für die Qualitätskategorien, die jeweils mittels entsprechender Kriterien beschrieben werden. Mit Kontextorientierung wird das Aufzeigen eines

Beziehungsnetzes zwischen dem Experiment und inhaltlichen bzw. lebensweltlichen Aspekten des Unterrichts verbunden. Reflexivität berücksichtigt den Aspekt, dass das Wissen und die Expertise den Erkenntnisprozess beeinflussen und der Prozess der Erkenntnisgewinnung seinerseits das Wissen und die Expertise ausschärfen. Unter theoriegeleiteter Aktivität werden Lerngelegenheiten subsumiert, die naturwissenschaftliche Ideen, Vorstellungen und Konzepte transportieren und ein entsprechendes Denken fördern. Dieses Erhebungsinstrument ist für die Analyse von Experimentalstunden in Deutschland, Finnland und der Schweiz eingesetzt, aber soweit ersichtlich, nie validiert worden.

Im Projekt ExMo (Vermittlungs- und Beurteilungskompetenzen zum Experimentieren: Modellierung, Validierung und Messinstrumententwicklung) ist ein Instrument zur Erfassung der Planungs-, Analyse- und Diagnosekompetenz angehender Lehrpersonen in Bezug auf die Gestaltung von Experimentalunterricht entwickelt worden (Hasse et al., 2014). Das Instrument fasst den Experimentierprozess als Zusammenspiel unterscheidbarer Teilprozesse im Sinne eines idealisierten, hypothetisch-deduktiven Erkenntnisprozesses auf und beschreibt in Anlehnung an die drei Hauptphasen (Hypothesenformulierung, Planung des Experiments, Datenanalyse) die entsprechenden Lehrkompetenzen (Hasse et al., 2014, S. 198). Dabei werden die Planungs-, Analyse- und Diagnosekompetenz mit Hilfe von je drei Facetten zu den drei Hauptphasen beschrieben. Die Planungskompetenz wird beschrieben mit den Facetten (a) wie Hypothesen aufzustellen sind, (b) wie ein Experiment geplant wird und (c) wie die erhobenen Daten analysiert werden. Das Instrument ist im Rahmen einer Pilotstudie eingesetzt, aber statistisch noch nicht geprüft worden.

Für die Entwicklung naturwissenschaftlicher Kompetenz wird die Gestaltung des Unterrichts als entscheidender Bedingungsfaktor eingestuft (Seidel et al., 2006). Entsprechend bedarf es einer grundlegenden Wissensbasis, insbesondere auch einem fachdidaktischen Wissen zum experimentellen Handeln im Sinne der Erkenntnisgewinnung. Beschreibungen der dazu notwendigen Lehrerexpertise und davon abgeleitete Messinstrumente sind allerdings, wie vorangehend beschrieben, insgesamt rar (Bögeholz et al., 2016; Hasse et al., 2014, S. 195). Wissen lässt sich gemäss Baer et al. jedoch auf der Basis von Kriterien qualitätsvollen Unterrichts in Unterrichtsplanungen erfassen (Baer et al., 2011). Unterrichtsplanungen bieten demnach eine geeignete Möglichkeit, Einblick in das handlungsvorbereitende Wissen und die Kompetenzen der Studierenden zu erhalten (Terbrügge, 2001; Tillema, 2009). Das nächste Kapitel befasst sich vor diesem Hintergrund mit dem Thema der Unterrichtsplanungen als Möglichkeit, Einblick in die Kompetenzen der Studierenden zu erhalten.

2.8 Unterrichtsplanungen – Einblick in die Kompetenzen der Studierenden

Das Planen von Unterricht gehört zusammen mit seiner Durchführung und Auswertung zu den auch in den Standards festgelegten (cf. Kap. I, 2.4.3) beruflichen Kernaufgaben und Tätigkeiten von Lehrpersonen (Blömeke, Kaiser, et al., 2008; Greiten, 2015; Kunter & Baumert, 2006; Seel, 2011). Der Planungsprozess, der als Form didaktischen Denkens und Handelns aufgefasst werden kann (Wernke & Zierer, 2017), wird dabei in der Fachliteratur aufgrund des jeweils gewählten Fokus unterschiedlich ausgelegt und bezeichnet. Heymann (2007) unterscheidet zwischen der Unterrichtsvorbereitung und der Unterrichtsplanung. In die Unterrichtsvorbereitung schliesst er alle Tätigkeiten ein, die im Dienst der Vorbereitung eines späteren Unterrichts stehen. Mit Unterrichtsplanung bezeichnet er nur das eigentliche, dem Unterricht vorgelagerte konkrete Planungshandeln. Andere Autoren verwenden diese beiden Begriffe aufgrund der schweren Abgrenzbarkeit und Zugänglichkeit des Forschungsgegenstandes auch synonym (e.g. Gassmann, 2013; Seel, 2011). Letzteres deckt sich mit dem Verständnis der vorliegenden Studie und soll näher erläutert werden.

2.8.1 Unterrichtsplanung – ein antizipierendes Probehandeln für den Kompetenzaufbau von (angehenden) Lehrpersonen

Die Unterrichtsplanung im Sinne einer vorgelagerten Tätigkeit, stellt ein „antizipierendes, gedankliches Probehandeln“ dar, das „als Prozess der Informationsverarbeitung auf einen Handlungsentwurf“, d.h. ein Produkt, abzielt (Kiper, 2011 S. 132). Das Planungshandeln kann somit als Problemlösungs- und Entscheidungsprozess aufgefasst werden (Terbrügge, 2001). Demgemäss handelt es sich um einen „kognitiven Vorgang [...], der selbst wieder kognitive Inhalte, nämlich Wissen, Vorstellungen, Erinnerungen zum Gegenstand hat und dessen Produkt, der Plan nämlich, selbst wieder Wissen ist“ (Bromme & Seeger, 1979, S. 4). Pläne sind demnach domänenspezifische Wissensbestandteile in handlungsleitender Funktion, wobei auch unbewusste Anteile darin enthalten sind, die in erheblichem Mass von Gefühlen und Bedürfnissen gesteuert werden (Seel, 2011; Stender et al., 2017; Terbrügge, 2001). Das Planungshandeln erfordert damit „die Ausdifferenzierung, Integration und Transformation“ der verschiedenen Bereiche des Professionswissens, insbesondere der drei Wissensdimensionen gemäss Shulman (1987) (Seel, 2011, S. 31). Ausserdem setzt es eine Bestimmung der für das Gelingen von Unterricht wichtigen Faktoren voraus (Kiper & Mischke, 2009). Im Planungshandeln lassen sich daher in gewisser Weise Theorie und Praxis zusammenführen (Causton-Theoharis, Theoharis & Trezek, 2008). Entsprechend werden Fähigkeiten im Bereich der Unterrichtsplanung als integraler, aber bisher wenig

beachteter Bestandteil professioneller Handlungskompetenz aufgefasst (von Aufschnaiter & Blömeke, 2010). Zierer (2015) spricht in diesem Zusammenhang nicht von Handlungskompetenz, sondern von Expertise und erweitert damit den aus seiner Sicht sehr stark auf Wissen und Fähigkeiten reduzierten Begriff der Handlungskompetenz mit Kriterien wie Volition und Urteilsfähigkeit.

In der Metastudie von Hattie & Yates (2015) wird das Planungshandeln als eigentliche Handlung der Lehrperson als Schlüssel für erfolgreiches Unterrichten betont. Unterrichtsplanungen und Entscheidungen, die im Planungsprozess getroffen werden, begünstigen das spätere Verhalten der Lehrperson im Unterricht und beeinflussen auch den Unterrichtsverlauf (Clark & Peterson, 1986; Frudden, 1984; Seel, 1997; Shavelson, 1983; Weingarten & van Ackeren, 2017). Eine „gute“ Planung kann daher hilfreich sein, um vorab entsprechende Faktoren zu fokussieren (Zierer, Werner, et al., 2015) und scheint eine wichtige Voraussetzung für angemessene Anforderungen an die Lernenden zu sein (Lipowsky, 2015). Für die Entwicklung kompetenten Handelns spielt demnach das Erstellen von Unterrichtsplanungen eine wesentliche Rolle (Baylor, 2002). Unterrichtsplanungen sind somit Ausdruck transformierter Wissensbestände in Form von Handlungsplänen oder Skripts. Der nächste Abschnitt geht auf diesen Aspekt näher ein.

2.8.2 Unterrichtsplanung – ein Ausdruck von Wissensbeständen in Form von Handlungsplänen oder Skripts

Für die Gestaltung von Unterricht greifen Lehrpersonen auf strukturierte Wissensbestände in Form von Plänen, Skripts und Routinen zurück (Pauli & Reusser, 2003). Die Unterrichtsplanung übernimmt damit die Funktion eines Katalysators bei der Transformation von Professionswissen in Handlungsskripte (Stender et al., 2015). Das kognitionspsychologische Konzept des Skripts, das auf Schank und Abelson zurückgeht, wird beschrieben als „predetermined, stereotyped sequence of actions that defines a wellknown situation“ (Schank & Abelson, 1977). Dieses Konzept wird in der Literatur sehr unterschiedlich aufgefasst, sei es als länderspezifisches Verlaufsmuster von Unterricht oder aber als spezifischer Unterrichtsablauf zur Vermittlung bestimmter Lerninhalte bzw. zur Erreichung bestimmter Zielsetzungen (cf. Pauli & Reusser, 2003). In Bezug auf den spezifischen Unterrichtsablauf sind jedoch im Konzept der Skripts die Ebenen des Unterrichts nicht geklärt, d.h. der Bezug zur Oberflächen- bzw. Tiefenstruktur. Gemäss Pauli und Reusser (2003) sind stereotype Abläufe eher auf die Inszenierung und damit die Oberfläche des Unterrichts zu beziehen, während die Ebene der Tiefenstruktur planvolle, auf Ziele ausgerichtete und bewusste Entscheidungen in Form von Plänen verlangt. Stender, Brückmann und Neumann (2015) gehen aber davon aus, dass Handlungspläne durch mehrfache Verbesserung in

Handlungsskripte überführt werden können. Vor diesem Hintergrund hat Stender (2014) ein theoretisches, fachspezifisches Transformationsmodell der Unterrichtsplanung entwickelt, das die Transformation deklarativen Wissens in Handlungsskripte konkretisiert und neben dem Professionswissen auch die affektiven Dispositionen berücksichtigt. Grundlage für dieses Modell bildet zum einen die kognitionspsychologische ACT-Theorie (*Adaptive Control of Thought*) von Anderson (1976) sowie der von Shavelson und Stern (1981) beschriebene Planungsprozess, der als zirkulärer kognitiver Entscheidungsablauf aufgefasst wird. Darin werden, ausgehend von Analysen der Vorbedingungen (z.B. Lernvoraussetzungen), in Selektions- und Integrationsprozessen Entscheidungen für Unterrichtsaktivitäten getroffen, die wiederum nach dem Unterricht reflektiert und modifiziert werden. Dieses empirisch geprüfte Transformationsmodell der Unterrichtsplanung ist in Bezug auf die Wissensdimensionen weiter ausgeschärft worden (Abb. 12). Die Grundlage für diese Weiterentwicklung bildet das Konsensmodell von Berry et al. (2015; cf. Kap. I, 2.5.2).

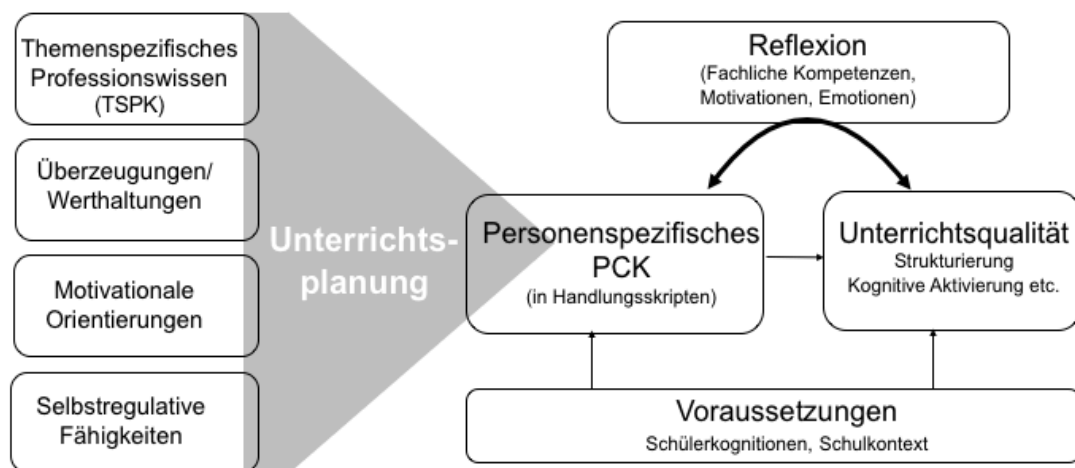


Abbildung 12: Transformationsmodell zum Planen von Unterricht (Stender et al., 2017, S. 1695)

Das adaptierte, fachspezifische Modell, das unter Berücksichtigung der affektiven Dispositionen die Transformation des themenspezifischen Professionswissens (TSPK) in Handlungsskripte und damit in ein personen-, themen- und kontextspezifisches PCK modelliert, ist erneut geprüft und durch die Ergebnisse empirisch gestützt worden (Stender et al., 2017). Die Ergebnisse der Untersuchung unterstreichen auch in diesem Modell die Bedeutung der affektiven Dispositionen, die den Einfluss des fachdidaktischen Wissens auf die Qualität der Handlungsskripts moderieren (ebd.). Unterrichtsplanungen können demzufolge als zentraler Prozess für die Interpretation und Kompilierung des Professionswissens angesehen werden (Stender, 2014, S. 35), insbesondere von themenspezifischem

Professionswissen (TSPK). Die Kompetenz des Planungshandelns manifestiert sich somit in der Qualität der entwickelten Handlungspläne. Diese lässt sich vor dem Hintergrund normativer Setzungen einschätzen und an der Güte des Plans, die sich daran bemisst, ob es gelingt das angestrebte Ziel zu erreichen (Kiper & Mischke, 2009). Es muss dabei aber immer berücksichtigt werden, dass der Planungsprozess vielschichtig ist und von den Akteuren meist nicht vollständig verbalisiert wird (Seel, 2011).

Für die iterative Weiterentwicklung und Festigung der Handlungspläne und damit auch für das personen-, themen- und kontextspezifische PCK wird der Reflexion eine wichtige Bedeutung beigemessen (Stender, 2014, S. 39). Diese Erkenntnis deckt sich mit Befunden von Baer et al. (2011), die darlegen, dass eine kontinuierliche Auseinandersetzung mit der Unterrichtsplanung das Wissen über guten Unterricht und damit die Unterrichtsqualität verbessert. Van der Valk & Broekman (1999) kommen in ihrer Studie zur Erkenntnis, dass angehende Lehrpersonen durch die Planung von Unterricht und einem anschließenden Gespräch darüber nicht nur ihr vorhandenes PCK zeigen, sondern dieses auch weiterentwickeln. Ähnliche Evidenzen werden ausserdem in Arbeiten berichtet (cf. Ruys, Keer & Aelterman, 2012). Mutton, Hagger & Burn (2011) äussern sich in ihrer Studie dazu wie folgt: „[...] ongoing learning about planning can be a powerful vehicle for ongoing learning about teaching as a whole.“ Im Kontext der Lehrpersonenbildung und der Professionalisierungsforschung stellt das Planungshandeln somit eine wichtige Schlüsselstelle des Unterrichtens dar (Zierer, Werner & Wernke, 2015a). Das Erstellen von Unterrichtsplanungen spielt für die Entwicklung kompetenten Handelns und in diesem Zusammenhang auch von PCK eine wesentliche Rolle (Ball & Bass, 2000; Baylor, 2002). Unterrichtsplanungen bieten daher einen geeigneten Zugang, um Einblick in die kognitiven Aktivitäten und die Kompetenzen der Studierenden zu erhalten (Davies & Rogers, 2000; König, 2015; Seel, 1997; Tillema, 2009). Ruys, Keer und Aelterman (2012) stufen unterrichtliche Planungen in der Kompetenzbeurteilung von Lehrpersonen gar als Bereicherung ein.

Trotz der anerkannten Bedeutung der Unterrichtsplanung erfährt das Planungshandeln von (angehenden) Lehrpersonen wie auch die Entwicklung der Fähigkeit zur Unterrichtsplanung in der Forschung vergleichsweise wenig Aufmerksamkeit. Es ist daher nach wie vor ein eher randständiger und vernachlässigter Forschungsbereich (Blömeke & König, 2011; Gassmann, 2013; Seel, 2011, S. 32). Der Erkenntnisstand ist insgesamt lückenhaft. Diese Sachlage soll auf der Grundlage empirischer Befunde zuerst allgemein, dann mit dem Fokus auf der Analyse von Unterrichtsplanungen zum naturwissenschaftlichen Unterricht, im nächsten Abschnitt näher aufgezeigt werden.

2.8.3 Unterrichtsplanungen – Vielfalt der Herangehensweisen bei der empirischen Untersuchung

Befunde zum ausbildungsbezogenen und alltäglichen Planungsprozess sowie zur Expertise in Unterrichtsplanungen sind u.a. in Arbeiten von Clark und Peterson (1986), Haas (1998), Seel (1997, 2011) und Stender (2014) aufgeführt. Darunter sind auch Studien zu Unterrichtsplanungen naturwissenschaftlicher Themen zu finden (Duschl & Wright, 1989; Haas, 1998; So, 1997; Stender, 2014; Stender et al., 2017; Terbrügge, 2001). Eine systematische Zusammenstellung der Befunde wird allerdings u.a. durch die unterschiedlichen Stichproben und vielfältigen Forschungsfragen dieser Studien, wie auch wegen der zahlreichen forschungsmethodischen Zugänge erschwert. Ausserdem bestehen Differenzen in den jeweils den Studien zugrunde liegenden Professionalisierungstheorien und den länderspezifischen Lehrpersonenbildungssystemen, was die Vergleichbarkeit einschränkt (Seel, 2011).

Bei den Stichproben zeigt sich, dass der Forschungsschwerpunkt von Studien aus dem englischsprachigen Raum vielfach auf Unterrichtsplanungen zu verschiedenen Fächern in der Primarstufe liegt, während er bei deutschsprachigen Studien mehrheitlich auf Unterrichtsplanungen für die Sekundarstufe I liegt. Viele dieser Studien betreffen vor allem ausgebildete Lehrpersonen. Untersuchungen angehender Lehrpersonen sind vergleichsweise selten. Insgesamt lässt sich aus all den Studien kein einheitliches Konzept für die Erfassung Planungstätigkeit bzw. Planungskompetenz ableiten (König, Buchholtz & Dohmen, 2015; cf. für eine Übersicht Stender, 2014, S. 87ff.). Der Grund dafür liegt in den verschiedenen Forschungsanliegen und den damit verbundenen unterschiedlichen Forschungsfragen. Diese reichen von Fragen u.a.:

- zu externalen Planungsgründen
- zu Typen von Planungen
- zum Aufwand und zu den verwendeten Unterlagen bei der Planung
- zu den Planungsprozessen bzw. -schritten und den thematischen Elementen von Planungsmodellen und -konzepten, die darin einbezogen werden
- zu möglichen Faktoren, welche die Unterrichtsplanungen beeinflussen (e.g. Borko, Roberts & Shavelson, 2008).

Die Liste zeigt auch, dass Fragestellungen, die den Zusammenhang zwischen Unterrichtsplanung und konkreter Unterrichtsdurchführung berücksichtigen, bisher, mit Ausnahme von drei unlängst erschienenen Studien von Gess-Newsome et al. (2017), König et al. (2015) und Anthofer (2017), eher vernachlässigt worden sind (Seel, 2011, S. 32). Eine Folge der unterschiedlichen Untersuchungsfokusse sind auch die verschiedenen methodischen Herangehensweisen und Untersuchungsdesigns. So erfolgten die Untersuchungen im Rahmen des Planungshandelns indirekt (z.B. Interviews, *stimulated recall*, Lautes Denken,

Fragebogen, Vignetten, etc.) oder direkt (Beobachtung, Dokumentenanalyse) (Borko, Roberts & Shavelson, 2008; Clark & Peterson, 1986; Gassmann, 2013; Haas, 1998, 2005; Zierer, Wernke & Werner, 2015b). Die Planungskompetenz im Sinne einer kognitiven Disposition zur Planung, die sich in spezifischen Situationen zeigt, wird damit im Fall der indirekten Vorgehensweise vielfach „als berufspraktisches Wissen über die Unterrichtsplanung und nicht als kontextbezogenes Planungshandeln“ erfasst (Seel, 2011, S. 34). Situative Bedingungen, die bei der Unterrichtsplanung eine wesentliche Rolle spielen, werden demzufolge nicht immer in den Forschungsvorhaben berücksichtigt (König et al., 2015). Daher sind Studien vergleichsweise selten, die den Fokus auf eine standardisierte Analyse konkret ausgearbeiteter Planungsunterlagen legen und zu diesem Zweck ein Kodiermanual mit Indikatoren zur Rekonstruktion und Quantifizierung der Planungsentscheidungen entwickelt haben. Zu erwähnen sind hier einige wichtige Arbeiten, meist neueren Datums:

- mit ausgewähltem, fachunspezifischem Fokus des realisierten Planungshandelns (Haas, 2005; König et al., 2015; Ruys, Keer & Aeltermann, 2012; Spooner, Baker, Harris, Ahlgrim - Delzell & Browder, 2007; Weingarten & van Ackeren, 2017; Zierer, Werner & Wernke, 2015a)
- mit ausgewähltem, fachspezifischem Fokus (hier Naturwissenschaften & Technik) des realisierten Planungshandelns (Gess-Newsome et al., 2017; Goldston, Dantzler, Day & Webb, 2013; Hacker & Sova, 1998; Jacobs, Martin & Otieno, 2008; Ozogul, Olina & Sullivan, 2008; van der Valk & Broekman, 1999).

Die Mehrheit der hier erwähnten Studien untersucht Unterrichtsplanungen von angehenden Lehrpersonen. Die gewählten Untersuchungsfokusse sind in diesen Studien ebenfalls verschieden, was sich in der Ausgestaltung der jeweiligen Erhebungsinstrumente zeigt. Die Instrumente für fächerübergreifende Analysen von Unterrichtsplanungen sind dabei z.B. ausgerichtet auf beobachtbare Bereiche wie «Inhalte und deren Strukturierung» (Zierer, Werner & Wernke, 2015a) und «Kompetenzorientierte Gestaltung von Lernangeboten» (Weingarten & van Ackeren, 2017). Oder aber sie stützen sich eher auf einzelne, ausgewählte Qualitätsmerkmale wie «Didaktische Adaptivität von Aufgaben» (König et al., 2015), «Individualisierung» (Spooner et al., 2007) und «Kooperatives Lernen» (Ruys et al., 2012).

Die Instrumente für fachspezifische Analysen fokussieren demgegenüber die in den Planungen angelegten fachspezifischen Lernprozesse zu jeweils unterschiedlichen Themen. Ozogul, Olina und Sullivan (2008) haben beispielsweise ein Kodiermanual zur Erfassung geplanter Lernprozesse zum Thema Technik entwickelt. Das Manual besteht aus vier Kategorien («Ziele», «Fachinhalt & Material», «Ablauf der Lektion» und «Evaluation»), die ihrerseits durch 15 Items bzw. Indikatoren beschrieben werden. Das von Jacobs, Martin

und Otieno (2008) erarbeitete Ratinginstrument zur Beobachtung von Planungen naturwissenschaftlichen Unterrichts umfasst ebenfalls 4 Kategorien mit insgesamt 34 Indikatoren. Die modellierten Kategorien sind «Abstimmung mit Standards und Erkenntnissen aus der Forschung», «Kognitive und metakognitive Aspekte der Konzeption und Umsetzung der Lektion», «Soziokulturelle und affektive Aspekte der Konzeption und Umsetzung der Lektion» sowie «Darstellungen und Anwendungen der Praktiken in der Wissenschaft (Experimentieren *Hands-on*, *Nature of Science*, Umgang mit Fehlern etc.)». Der Prozess der Erkenntnisgewinnung wird in diesem Instrument partiell berücksichtigt, aber nicht im Sinne eines Fachinhaltes. Im Gegensatz dazu steht in der Arbeit von Goldston et al. (2013) die Planungskompetenz zum Prozess der Erkenntnisgewinnung im Vordergrund. Das Instrument ist u.a. in Anlehnung an ein Manual für die Beobachtung von Unterrichtsstunden zur Erkenntnisgewinnung (Bodzin & Beerer, 2003) entwickelt worden. Es beinhaltet 5 Hauptkategorien, mit denen ein Lernprozess bzw. ein Lernzyklus zum experimentellen Handeln beschrieben und erfasst wird («engage», «explore», «explain», «elaborate», and «evaluate») sowie zwei übergeordnete Kategorien zu Kriterien wie Zielen, Materialien, Zeit, Individualisierung. Das Instrument stellt keinen Bezug zu den Wissensdimensionen von Shulman (1987) her und wurde auf der Basis der erhobenen Daten mit Hilfe einer explorativen Faktorenanalyse strukturell validiert. Zu erwähnen ist ausserdem auch eine etwas ältere Studie von van der Valk und Broekman (1999). Der Fokus dieser Studie lag auf der Untersuchung von PCK. Unterrichtsplanungen zu Themen aus den Fächern Mathematik, Physik und Chemie sind hier hinsichtlich Shulmans (1987) zentraler Facetten von PCK, nämlich «Schülervorstellungen» und «Instruktionsstrategien» untersucht worden. Die für die Analyse generierten Kriterien wurden jedoch, soweit ersichtlich, nicht validiert. Im Rahmen einer aktuellen Interventionsstudie von Gess-Newsome et al. (2017) wurde zur Erfassung des fachdidaktischen Wissens (PCK) von Biologielehrpersonen der Sekundarstufe I ein Ratinginstrument entwickelt. Dieses Instrument ist mit Bezug zu ausgewählten biologischen Fachinhalten themenspezifisch angelegt worden und besteht aus drei Dimensionen. Die erste Dimension, PCK-CK, umfasst Kriterien zu fachlichen Aspekten, wie die Richtigkeit des Fachinhalts, Vernetzungen zwischen Fachinhalten, multiple Repräsentationsformen und Beispiele zum Fachinhalt, aber auch Kriterien zu *Nature of Science*. Die zweite Dimension, PCK-PK, beinhaltet Kriterien zu möglichen Lehr- und Lernstrategien, wie beispielsweise Strategien zur Erhebung von Präkonzepten von Schülerinnen und Schülern, aber auch mögliche Fördermassnahmen, mit denen die Schülerinnen und Schüler ihr Denken selber reflektieren. Die dritte Dimension, PCK-CxK, setzt sich zusammen aus Kriterien, mit denen das Verständnis zur Heterogenität der Schülerinnen und Schülern und sich daraus ableitende Instruktionsstrategien erhoben werden. Für die Validierung des Instruments und für erste Analysen lieferten Erläuterungen von Unterrichtsplanungen und

Interviews die Datengrundlage. Von den drei festgelegten Dimensionen des Ratinginstruments konnten mittels Faktorenanalyse insgesamt zwei bestätigt werden. PCK-CK erwies sich als distinkte Dimension, während die anderen beiden Dimensionen zu einer Dimension zusammengefasst und von der Autorenschaft mit PCK-PK bezeichnet wurde.

Die Befundlage zu standardisierten Analysen von Unterrichtsplanungen zeigt, dass aus den bisherigen Ergebnissen keine konsistente Beschreibung des Planungshandelns und damit der Planungskompetenz von Lehrpersonen entwickelt werden kann (Vogelsang & Riese, 2017; Weingarten & van Ackeren, 2017). Entsprechend bleibt unklar, wie sich die Planungskompetenz tatsächlich operationalisieren lässt (Zierer, Werner & Wernke, 2015a). Die bisher entwickelten Erhebungsinstrumente zur Erfassung fachunspezifischer und fachspezifischer Qualitätskriterien, mit deren Hilfe die Planungsperformanz ermittelt worden ist, unterscheiden sich in den gewählten Untersuchungsausschnitten erheblich. Das experimentelle Handeln im Sinne der Erkenntnisgewinnung ist bisher nur in der Studie von Goldston, Dantzer, Day und Webb (2013) thematisiert worden. Einen expliziten Bezug zu PCK und seiner Erfassung in Planungen findet sich nur in der Studie von van der Valk und Broekman (1999). Es fällt weiter auf, dass in keiner der Studien der Fokus auf didaktische Modelle oder verwendete Planungsraster gelegt, sondern das Planungshandeln allgemein untersucht wird (Zierer, Wernke & Werner, 2015b). Die Erstellung von Unterrichtsplanungen auf der Grundlage fach- bzw. allgemein-didaktischer Modelle stellt jedoch nach wie vor ein wichtiges Element der akademischen Lehrpersonenbildung dar (cf. Schüle, Besa & Arnold, 2017). Der nächste Abschnitt befasst sich daher mit verschiedenen (fach-)didaktischen Planungsmodellen, die in der Ausbildung angehender Lehrpersonen thematisiert und eingesetzt werden.

2.8.4 Unterrichtsplanung – (fach-)didaktische Planungsmodelle als theoretischer Rahmen und Grundlage für die Planung

Für die Planung von Unterricht bilden überfachliche, allgemeindidaktische Modelle einen normativ festgelegten theoretischen Rahmen (König et al., 2015; Zierer, 2012). Sie reduzieren und strukturieren die Komplexität des zu planenden Unterrichts und bieten eine fundierte Grundlage zur reflektierten und theoriegeleiteten Planung und Analyse von Unterricht (Werner et al., 2017). Trotz des Anspruchs einer Gesamtschau auf den Unterricht stellen Modelle aber stets nur einen Ausdruck einer ganz bestimmten Sichtweise dar (Gassmann, 2013). Entsprechend ist die Fülle an didaktischen Planungsmodellen gross (Arnold, Rakhkochkine & Mitter, 2010; Heimann, Otto & Schulz, 1965; Klafki, 1958, 1980; Kron, Jürgens & Standop, 2014; Meyer, 2009). Der Einsatz allgemeindidaktischer Modelle als Grundlage für die Gestaltung von Unterricht wird in der Literatur allerdings kontrovers

diskutiert. Während sie in gewissen Untersuchungen als unpraktikabel und für die Praxis als nicht relevant eingeschätzt werden (e.g. Bromme, 1981; Haas, 1998; Terbrügge, 2001), konnte in neueren Studien gezeigt werden, dass eine an allgemeindidaktischen Modellen orientierte Planung die Gestaltung eines qualitätsvollen Unterrichts massgeblich unterstützen kann (e.g. Zierer, Werner & Wernke, 2015a). Ungeachtet dieser Befunde stellt eine, an allgemeindidaktischen Modellen orientierte Unterrichtsplanung nach wie vor ein wichtiges Element akademischer Lehrpersonenbildung dar (Schüle et al., 2017). Aus diesem Grund sollen zentrale, in Standardwerken der Allgemeinen Didaktik vielfach rezipierte und in der akademischen Lehrpersonenbildung u.a. thematisierte Modelle nächstfolgend beleuchtet werden. Es sind dies zum einen die Planungsmodelle der bildungstheoretischen und kritisch-konstruktiven Didaktik, wie das Modell der «Didaktischen Analyse» (Klafki, 1958) sowie das «Perspektivenschema zur Unterrichtsplanung» (Klafki, 1980), zum anderen aber auch die Modelle der lehr- und lerntheoretischen Didaktik, wie das «Berlinermodell» (Heimann et al., 1965) und das «Hamburgermodell» (Schulz, 1980) (cf. Arnold & Koch-Priewe, 2010). Die Modelle beider didaktischer Traditionen spielen für das fachdidaktische Modell der «Didaktischen Rekonstruktion», das anschliessend näher vorgestellt wird und Gegenstand der vorliegenden Untersuchung ist, eine wesentliche Rolle.

I. Modelle in der Tradition der bildungstheoretischen Didaktik

Die «Didaktische Analyse» (Klafki, 1958) und das «Perspektivenschema zur Unterrichtsplanung» (Klafki, 1980), zwei Modelle des bildungstheoretischen Ansatzes, verstehen „Unterricht als Prozess der bildenden Begegnung zwischen ausgewählten geeigneten Bildungsgütern und der nachwachsenden Generation“ (Terhart, 2008, S. 15). Im Fokus dieser beiden Modelle stehen die Inhalts- und Zieldimension des Unterrichts (Schüle et al., 2017). Im Kern geht es dabei um die Auswahl, Anordnung und Explikation unterrichtlicher Inhalte. Diese sollen kategorial bildend, d.h. elementar im Hinblick auf die Sache, fundamental im Hinblick auf die Schülerinnen und Schüler sein und exemplarisch erarbeitet werden (cf. Standop & Jürgens, 2015). Im Planungsmodell der «Didaktischen Analyse» (Klafki, 1958) erfolgt diese Auswahl aufgrund vorgängiger Lehrplanentscheidungen sowie aufgrund des Bildungsgehalts und der lebensweltlichen Relevanz der Themen. Damit schreibt Klafki dem Fachwissen der Lehrperson vor dem Hintergrund seines jeweiligen Bildungsgehalts eine grosse Bedeutung für den Unterricht zu (Brückmann, 2009). Die starke Fokussierung auf das kategoriale Bildungsverständnis und die im Modell angelegte, einseitige Ausrichtung auf die Notwendigkeiten und Anforderungen der Gesellschaft haben neben Zustimmung auch Kritik am Modell der «Didaktischen Analyse» ausgelöst (Arnold & Koch-Priewe, 2010; Standop & Jürgens, 2015). Allgemein ist auch der im Modell fehlende Bezug zur unterrichtlichen Wirklichkeit bemängelt worden (Standop & Jürgens, 2015).

Klafki hat deshalb das Modell der «Didaktischen Analyse» im Sinne der konstruktivistisch-kritischen Didaktik weiterentwickelt und daraus ein weiteres Planungsmodell, das «Perspektivenschema zur Unterrichtsplanung» (Klafki, 1980) erarbeitet. Das Modell orientiert sich nach wie vor am kategorialen Bildungsbegriff. Zentraler Orientierungspunkt dieses neueren Modells bildet aber die Entfaltung der Persönlichkeit junger Menschen. Richtungsweisend für die Wahl der Inhalte sind daher primär übergeordnete Zielentscheidungen wie die Selbstbestimmungs-, Mitbestimmungs- und Solidaritätsfähigkeit. Das bedeutet, dass sich auf der Ebene der Lehrpläne die Aufgabe stellt, zwischen den jeweils aktuellen Interessen, Erfahrungen und Bedürfnissen der Lernenden sowie ihren alltagsweltlichen Problemen und den Ansprüchen ihrer zukünftigen gesellschaftlichen und individuellen Aufgaben zu vermitteln (Tulodziecki, Herzig & Blömeke, 2017). Grundlegende Änderungen, die in diesem Modell vorgenommen worden sind, beziehen sich auf die Prozesse der Vorbereitung und Analyse von Unterricht (Standop & Jürgens, 2015). Leitende Prinzipien für die Themenwahl sind demzufolge nicht mehr die Themen aus den Schulfächern, sondern epochale Schlüsselprobleme für das gesellschaftliche Zusammenleben, mit deren Hilfe die übergeordneten Ziele erreicht werden können. Bildung wird so gesehen zum Interaktionsprozess und somit auch zum sozialen Prozess (Klafki, 1999, S. 15f.).

In beiden Planungsmodellen Klafkis werden Fragen zur methodischen Ausgestaltung und Sequenzierung der Lernprozesse, Fragen zur medialen Unterstützung, wie auch Fragen zur Analyse der Lernfortschritte etc. erst in einem zweiten Schritt geklärt (Terhart, 2008). Dies hat erneut Kritik ausgelöst (Arnold & Koch-Priewe, 2010; Gassmann, 2013). Gemäss Straka und Macke (1979) ist der Grund für die Kritik darin zu suchen, dass die Modelle nach wie vor von den Aussenbedingungen, wie dem Bildungsbegriff der jeweiligen Gesellschaft, ausgehen und mit den gestellten didaktischen Fragen zu wenig gezielt auf die Ebene der Lernenden vordringen. Es bleibt daher ungeklärt, nach welchen Prinzipien die Auswahl der Unterrichtsinhalte mit den Unterrichts- und Lernmethoden sowie der Strukturierung der Lehr- und Lernsequenzen verschränkt werden sollen (Straka & Macke, 1979, S. 112). Auch Heimann hat den von Klafki postulierten Bildungsbegriff für eine praktikable Didaktik als ungeeignet erachtet (Heimann, 1962). Nach seiner Auffassung sollte an Stelle des Bildungsbegriffs der Lernbegriff stärker fokussiert werden.

Auf der Grundlage dieses didaktischen Verständnisses sind deshalb zwei Planungsmodelle als Gegenentwurf zur bildungstheoretischen Didaktik entwickelt worden. Diese werden in der Folge vorgestellt.

II. Modelle in der Tradition der lehr- und lerntheoretischen Didaktik

Das «Berlinermodell» (Heimann et al., 1965) und das «Hamburgermodell» (Schulz, 1980), zwei Modelle der lehr- und lerntheoretischen Didaktik, setzen die Planung und Analyse von Unterricht in den Mittelpunkt ihrer Betrachtungen. Dabei geht es um die „zweckrationale und erfolgskontrollierte Organisation von Lehr-Lern-Prozessen“ auf der Basis empirischer Forschung (Kron et al., 2014; Terhart, 2008, S. 17).

Das 1965 von Heimann et al. entwickelte «Berlinermodell» stellt einen Versuch dar, alle im Unterricht wirksamen Faktoren deskriptiv und wertneutral zu erfassen (Wiater, 2009, S. 506). Dieses sogenannte Strukturmomentenmodell basiert auf einem lehrtheoretischen Analysekonzept, das in eine Planungsstruktur übertragen worden ist. Es beinhaltet zum einen eine Strukturanalyse, die sich aus der Analyse von sechs Planungsbereichen zusammensetzt, wovon vier Entscheidungsfelder (Intentionalität, Thematik, Methodik, Medienwahl) sind und zwei Bedingungsfelder (anthropologisch-psychologische und soziokulturelle Voraussetzungen), die alle in Wechselwirkung zueinanderstehen. Zum anderen enthält es auch eine Faktorenanalyse. Im Mittelpunkt dieser Analyse stehen Faktoren aus dem Umfeld, welche die Entscheidungen beeinflussen können oder beeinflussen (Normenkritik, Faktenbeurteilung, Formenanalyse). Für den Prozess der Unterrichtsplanung sieht das Modell ein zweiphasiges Vorgehen vor. Im Rahmen der Strukturanalyse werden die Entscheidungs- und Bedingungsfelder mit ihren Wechselwirkungen geprüft (Reflexionsstufe I) und daran anschliessend folgt die Faktorenanalyse (Reflexionsstufe II). Eine weitere Grundlage dieses Modells bilden ausserdem drei zentrale Prinzipien (Interdependenz, Variabilität, Kontrollierbarkeit), die bei der Planung berücksichtigt werden sollen. Schulz (1965) begründet diese Prinzipien wie folgt:

Die Auffassung der Didaktik als eine Strukturtheorie des Lehrens und Unterrichtens führt zum Prinzip der *Interdependenz* der Unterricht konstituierenden Momente. Die anthropologischen Bedingungen didaktischen Denkens in der Gegenwart fordern *Variabilität* des Unterrichts, die Möglichkeit der Mitsteuerung des Unterrichtsprozesses durch die Lernenden. Die empirische Selbstkontrolle didaktischen Denkens drückt sich im Planungsprinzip größtmöglicher *Kontrollierbarkeit* des Unterrichtserfolges auf. (Heimann et al., 1965, S. 44 f.)

Der unterrichtsanalytische Ansatz von Heimann et al. (1965) hat die deutschsprachige didaktische Diskussion stark beeinflusst (e.g. Blankertz, 1969/2000). Dennoch ist dieser Ansatz wegen seines Postulats der Wertfreiheit unterrichtsrelevanter Faktoren, seiner zu starken Ausrichtung auf die Perspektive der Lehrpersonen und seines Fokus auf der Unterrichtsanalyse auch kritisiert worden (cf. Straka & Macke, 2009; Tulodziecki et al., 2017).

Ein Teil dieser Kritikpunkte hat Schulz aufgenommen und daraus später einen eigenständigen Modellentwurf, das «Hamburgermodell» (Schulz, 1980), entwickelt. Darin rückt er ab vom Primat der Wertfreiheit und stellt „die Unterrichtsgestaltung sowie die Gestaltung des Schullebens als aktive Übung in Emanzipation“ in den Mittelpunkt der Betrachtungen (Hallitzky, Marchand & Seibert, 2013). Entsprechend ist das Modell zu einem dynamischen Prozessmodell weiterentwickelt worden. Im Zentrum dieses Handlungsmodells steht die Planung von Unterricht, die durch vier, sich wechselseitig bedingende Handlungsmomente (Unterrichtsziele, Ausgangslage, Vermittlungsvariablen, Erfolgskontrolle) beschrieben wird. Dabei wird der Interaktion der Lehrperson und der Schülerinnen und Schüler in gleicher Weise Rechnung getragen. Unterricht wird damit nicht mehr als ein Vorgang der Information, sondern als ein Vorgang der Interaktion verstanden, der in Abhängigkeit zum soziokulturellen Bedingungsfeld steht und demzufolge gesellschaftlichen sowie institutionellen Einflüssen ausgesetzt ist (Gassmann, 2013). Die so stärker erfolgende Berücksichtigung der Lernenden bedingt eine Öffnung des Unterrichts. Diese erschöpft sich allerdings im lerntheoretischen Planungsmodell, denn es werden keine konkreten Hilfen für die Durchführung und Gestaltung von Unterricht gegeben (Hallitzky et al., 2013).

Neben der Praktikabilität wird von Seiten der empirischen Lehr- und Lernforschung und den Fachdidaktiken immer wieder der Vorwurf geäußert, dass Lehren und Lernen stets auch an einen Unterrichtsgegenstand gebunden sein sollte (Klieme et al., 2003, S. 24ff.; Wilhelm, 2007; cf. Kap. I, 2.4.3). Entsprechend sind die allgemeindidaktischen Planungsmodelle auch hinsichtlich dieses Aspektes kritisiert worden. Vor diesem Hintergrund ist das Modell der «Didaktischen Rekonstruktion» entwickelt worden, ein fachdidaktisches Modell für die Unterrichtsplanung der Lehrpersonen, aber auch ein Rahmen für fachdidaktische Forschungs- und Entwicklungsprojekte (Kattmann, Duit, Gropengiesser & Komorek, 1997). Das in der Fachdidaktik etablierte Modell wird in der Folge vorgestellt.

III. Die «Didaktische Rekonstruktion» – ein Modell der fachdidaktischen Tradition

Das Modell der «Didaktischen Rekonstruktion» ist auf der Basis von bildungstheoretischen, allgemeindidaktischen und fachdidaktischen Traditionen zur Unterrichtsplanung und Curriculumsentwicklung erarbeitet worden (Kattmann et al., 1997). Es stützt sich auf die „klassischen“ deutschen Ansätze zur Didaktik (Komorek & Kattmann, 2008). Im Modell werden essenzielle Aspekte der «Didaktischen Analyse» von Klafki (1958) sowie des Strukturmomentenmodells der „Berlinschule“ von Heimann, Otto und Schulz (1965) berücksichtigt (Duit, 2004).

Entsprechend baut das Modell auf der Idee der Sachanalyse unter didaktischem Aspekt auf sowie auf der Sichtweise der grundlegenden Interdependenz aller den Unterricht bestimmenden Variablen (Kattmann, Duit, Gropengiesser & Komorek, 1997).

Zwei zentrale Annahmen liegen dem Modell zugrunde (van Dijk & Kattmann, 2007):

- Lernen erfolgt immer in einem situationalen Kontext, der nicht durch das wissenschaftliche Wissen vorgegeben ist, sondern hergestellt werden muss.
- Das Bewusstsein der Schülervorstellungen beeinflusst die Rekonstruktion des entsprechenden Fachinhalts.

Das Modell gründet auf einer konstruktivistischen, epistemologischen Sichtweise (Duit, Treagust & Widodo, 2008; Widodo & Duit, 2004). Dabei sind zwei zentrale Facetten richtungsweisend. Zum einen geht das Modell von dem Verständnis aus, dass die Lernenden ihr Wissen selbst konstruieren und damit zum Referenzpunkt des Lernens werden. Zum anderen basiert es auf der Erkenntnis, dass das fachwissenschaftliche Wissen ebenfalls eine Konstruktion des Menschen darstellt, die als Konsens einer *Scientific Community* aufgefasst werden kann (Duit, 2007). Vor diesem Hintergrund werden im Modell die Perspektiven der Lernenden und die Sachstruktur des Lerngegenstandes durch geeignete didaktische Strukturierung zusammengeführt (Wilhelm & Kalcsics, 2017; cf. Abb. 13).

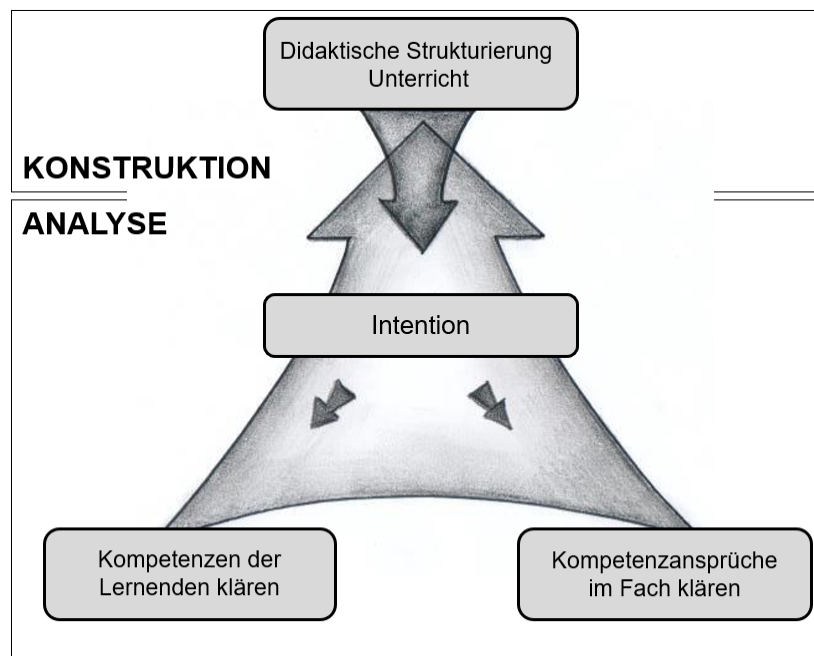


Abbildung 13: Modell der «Didaktischen Rekonstruktion» (verändert in Anlehnung an Kattmann et al., 1997 & Wilhelm & Kalcsics, 2017)

Dabei wird der wissenschaftliche Fachinhalt auf der Basis der Kompetenzen der Lernenden in Kombination mit der Analyse des Fachinhaltes und den Kompetenzansprüchen im Fach rekonstruiert. Die Kenntnisse zu den Kompetenzen der Lernenden (Schülervorstellungen, Interessen und Einstellungen, Denk-, Arbeits- und Handlungsweisen zu den Sachkonzepten) eröffnen einen neuen Blick auf die Sachstruktur der Wissenschaft und führen zu einem vertiefteren Verständnis des wissenschaftlichen Inhaltes (Duit, 2004; Wilhelm & Kalesics, 2017). Der wissenschaftliche Inhalt wird in Bezug auf zentrale Konzepte analysiert und unter Einbezug der Vermittlungsabsicht bzw. der Intention (Ziele) und der Schülervorstellungen in einen Kontext verpackt, der für die Schülerinnen und Schüler verständlich ist (van Dijk & Kattmann, 2007; Duit et al., 2012). Im Anschluss an diese «Analyse» erfolgt die «Konstruktion» und in diesem Zusammenhang die didaktische Strukturierung konkreter und möglichst bedeutsamer Lerngelegenheiten, in denen lernförderliche Korrespondenzen und voraussehbare Lernschwierigkeiten sichtbar werden (Kattmann et al., 1997, S. 12). Das Vorgehen ist dabei stets rekursiv. Die Sachstruktur der Wissenschaft wird mit Hilfe der «Didaktischen Rekonstruktion» in die Sachstruktur für den Unterricht transformiert. Dabei soll dieser Transformationsprozess nicht nur für die Inhalte der Fachwissenschaft, sondern gleichermassen auch für die naturwissenschaftlichen Denk- und Arbeitsweisen (*Scientific Inquiry*) sowie für die Natur der Naturwissenschaften (*Nature of Science*) erfolgen (Duit et al., 2012). Dieser Prozess der Rekonstruktion und Transformation kann aber keineswegs als Simplifizierung aufgefasst werden, da die fachlichen Konzepte und die Schülervorstellungen gleichwertig in die Analyse einbezogen werden und der Fachinhalt auf die Bedürfnisse der Lernenden ausgerichtet werden muss. Entsprechend unterscheidet sich die aus der Transformation entstandene Sachstruktur des Unterrichts klar von der Sachstruktur der Wissenschaft (Duit et al., 2012). Das Modell der «Didaktischen Rekonstruktion» ist gekennzeichnet durch seinen dualen Charakter. Es stellt einerseits ein Rahmenmodell für die naturwissenschaftsdidaktische Forschung und Entwicklung dar. Andererseits ist es ein Planungsmodell für die Praxis und wird als wesentliche Anleitung für die Planung von naturwissenschaftlichem Unterricht verwendet. Bislang ist das Modell nie empirisch getestet worden, es basiert aber auf einem breiten theoretischen Fundament und hat sich in der fachdidaktischen Forschung etabliert (Duit et al., 2012). Die Bestimmung der Qualitätsmerkmale von Handlungsplänen auf der Basis dieser normativen Vorgabe scheint daher für die Erstellung von Planungen für den naturwissenschaftlichen Unterricht naheliegend, insbesondere auch für die Ausbildung von (angehenden) Lehrpersonen. Vor diesem Hintergrund stellen auf der einen Seite die Kompetenzen der Lernenden und die Kompetenzansprüche im Fach wichtige Entscheidungsfelder für die Beschreibung entsprechender Qualitätsmerkmale dar. Auf der anderen Seite sind es auch die Ziele oder Intentionen sowie die didaktische Strukturierung der Lerngelegenheit, die zur Beschreibung der

Qualität mitbedacht werden sollten. Demgemäss sollten (angehende) Lehrpersonen über ein entsprechendes deklaratives fachdidaktisches Wissen zu diesen Entscheidungsfeldern verfügen. Vergleicht man die Themen der Entscheidungsfelder mit Shulmans Kernfacetten von PCK, so sind Übereinstimmungen in den Facetten Schülervorstellungen, wie auch der didaktischen Strukturierung von Lerngelegenheiten festzustellen.

Die Kompetenzansprüche im Fach wie auch die Intentionen werden von Shulman allerdings nicht als Facetten von PCK aufgefasst, von anderen Autoren hingegen schon (cf. Kap. I, 2.5.2). Damit zeigt sich, dass das Modell der «Didaktischen Rekonstruktion» wichtige Hinweise zu Entscheidungsfeldern liefert, die für die Planung von Bedeutung sind. Das Modell selbst kann aber keine Aussagen machen zu PCK von (angehenden) Lehrpersonen, das sich in den Unterrichtsplanungen, soweit beobachtbar, manifestiert. Weil das Wissen und die Einstellungen bei der Planung von Instruktionsprozessen eine entscheidende Rolle spielen, haben van Dijk und Kattmann (2007) auf der Grundlage der «Didaktischen Rekonstruktion» das ERTE-Modell entwickelt. Dieses Forschungsmodell soll im nächsten Abschnitt zur besseren Einordnung der vorliegenden Studie vorgestellt werden.

2.8.5 ERTE-Modell – ein fachdidaktisches Forschungsmodell zur Analyse von PCK (angehender) Lehrpersonen

Das ERTE-Modell (*educational reconstruction for teacher education*) ist mit dem Ziel entwickelt worden, die fachdidaktische Lehrpersonenbildung zu verbessern (Kizil & Kattmann, 2013, Abb. 14).

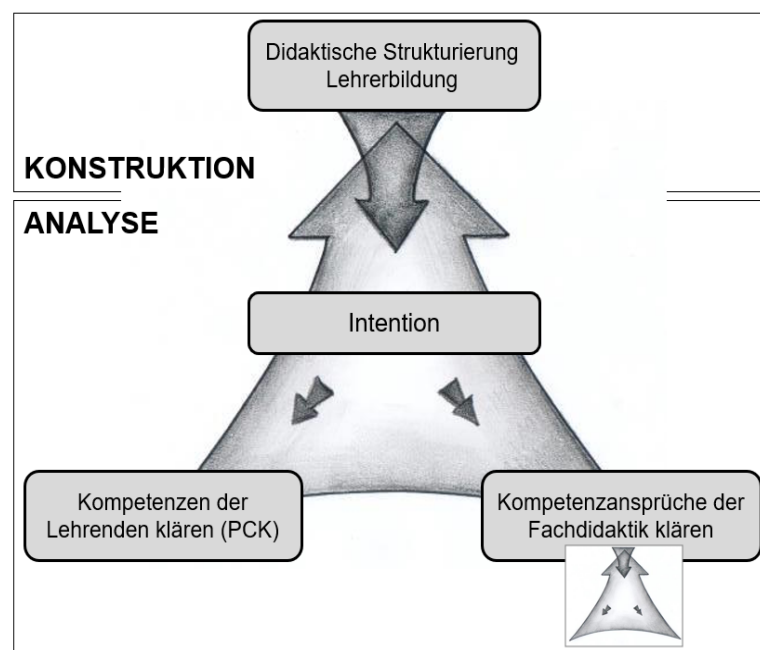


Abbildung 14: ERTE-Modell (verändert in Anlehnung an van Dijk & Kattmann, 2007 sowie Wilhelm & Kalcsics, 2017)

Es handelt es sich um ein Forschungsmodell, mit dessen Hilfe PCK von (angehenden) Lehrpersonen untersucht werden kann (van Dijk & Kattmann, 2007). Dem Modell liegt ebenfalls die Annahme zugrunde, dass die Inhalte der Lehrpersonenbildung für die Lehre rekonstruiert werden müssen (Kizil & Kattmann, 2013). Demgemäss bietet auch hier das Modell der «Didaktischen Rekonstruktion» einen theoretischen Rahmen (van Dijk & Kattmann, 2010). Dabei werden die Kompetenzen der Lehrenden, hier mit dem Fokus auf PCK, erfasst und mit den Kompetenzansprüchen der Fachdidaktik bezüglich fachlicher Klärung, Schülervorstellungen und didaktischer Strukturierung der Lerngelegenheit in Beziehung gesetzt (Kizil & Kattmann, 2013). Das in Unterrichtsplanungen beobachtbare PCK ermöglicht somit einen Einblick in die Qualität der geplanten Instruktionsprozesse (van Dijk & Kattmann, 2007). Insbesondere dem themenspezifischen PCK wird in der Literatur eine grosse Bedeutung beigemessen, da es Einblick in themenspezifische Instruktionsprozesse ermöglicht (Bucat, 2004). PCK wird gemäss Kizil und Kattmann (2013) über die (1) fachspezifischen Kenntnisse für das Lehren, (2) die Kenntnisse von Schülervorstellungen und (3) die Kenntnisse der Lehrenden von den verschiedenen Instruktionsformen des Unterrichtsthemas erfasst. Gemäss Ball und Bass (2000) können angehende Lehrpersonen über die Analyse und Gestaltung von Lerngelegenheiten ihr themenspezifisches, fachdidaktisches Wissen zeigen und entwickeln. Unklar bleibt jedoch, wie dieses fachdidaktische Wissen zur Planung von naturwissenschaftlichem Unterricht operationalisiert werden kann, insbesondere zum Experimentieren im Sinne der Erkenntnisgewinnung im Biologieunterricht. Entsprechende Instrumente sind, wie in Kapitel I, 2.8.3 gezeigt, rar. Bevor in Kapitel I, 2.10 die vorliegende Studie und insbesondere die Teilstudie I im Sinne eines Überblicks vorgestellt wird, folgt ein Fazit zum theoretischen und empirischen Hintergrund, auf dem die Untersuchung beruht.

2.9 Fazit

Mit der Einführung des kompetenzorientierten Lehrplans (D-EDK, 2014) und den zu erfüllenden Bildungsstandards (EDK, 2011) kommt der Vermittlung von Kompetenzen zum experimentellen Handeln im Sinne der Erkenntnisgewinnung eine grössere Bedeutung zu, insbesondere auch im Biologieunterricht. Die damit verbundene Auseinandersetzung mit dem Erkenntnisprozess, der als wissenschaftliche Problemlösung verstanden wird und sich prototypisch beim Experimentieren zeigt (Hammann, 2004; Mayer, 2007), wird auch als wichtiges Qualitätsmerkmal guten Biologieunterrichts erachtet (Neuhaus, 2007; Wüsten, 2010). Das notwendige Professionswissen und die damit verknüpfte professionelle Handlungskompetenz sollen in der Lehrerausbildung grundgelegt werden. Zur Einschätzung professioneller Handlungskompetenz genügen nach Oser et al. (2010) die nach Shulman

(1986/1987) definierten Wissensdimensionen nicht. Erst mit der Betrachtung der Handlungsebene wird die Handlungsrelevanz von erfasstem Professionswissen begründbar (Vogelsang & Reinhold, 2013a). Es kann aber vermutet werden, dass explizites Wissen schon im Unterrichtsplanungsprozess wirksam wird (Stender et al., 2015, 2017). Unterrichtsplanungen stellen daher ein zentrales Element auf dem Weg der Professionalisierung zum Lehrerberuf dar (John, 2006; Mutton, Hagger & Burn, 2011). Darüber hinaus eröffnen sie die Möglichkeit, Einblick in die Kompetenzen und in diesem Zusammenhang auch das handlungsvorbereitende Professionswissen der Studierenden zu erhalten (Baer et al., 2011; Stender, Brückmann & Neumann, 2017; Terbrügge, 2001; Tillema, 2009). Für die Ausübung der Lehrerverberuf wird das fachdidaktische Wissen als wichtige Grundlage eingestuft (Kind, 2009). Die in Kap. I, 2.5.2 dargestellte Synopse des Konsensmodells von Berry, Friedrichsen & Loughran (2015) und des Modells der Kompetenzmodellierung von Blömeke, Gustafsson & Shavelson (2015) stellt einen Versuch dar, die Dynamik der unterrichtlichen Tätigkeit von Lehrpersonen, auf der Basis des fachdidaktischen Wissens im Sinne der Wirkkette, möglichst ganzheitlich abzubilden. Bezugspunkt bildet dabei das themenspezifische Professionswissen, das im Rahmen des Planungsprozesses zu *PCK on action* bzw. *PCK in action* gemäss Stender, Brückmann & Neumann (2017, Kap. I, 2.8.2) transformiert wird. Für die Entwicklung fachdidaktischen Wissens kommt daher dem Planungshandeln eine wesentliche Bedeutung zu (Ball & Bass, 2000; Baylor, 2002). Beschreibungen der entsprechenden Expertise, insbesondere zum Planen von Unterricht zum experimentellen Handeln als Fachinhalt, sind derzeit jedoch rar. Auch sind bislang entsprechende Erhebungsinstrumente, mit deren Hilfe eine standardisierte Erfassung der Qualität der Planungsperformanz in Bezug auf das beobachtbare fachdidaktische Wissen erfolgt, selten und inkonsistent (Hasse et al., 2014; Vogelsang & Riese, 2017; Weingarten & van Ackeren, 2017). Die vorliegende Studie möchte diese Lücke schliessen. Sie legt den Fokus auf die Analyse von *PCK on action*, d.h. auf die Analyse deklarativen fachdidaktischen Wissens zum Planen von Unterricht zum experimentellen Handeln. Das nächste Kapitel konkretisiert die Teilstudie I des Dissertationsvorhabens im Sinne eines ersten Überblicks.

2.10 Überblick über Teilstudie I der vorliegenden Dissertation

Die vorliegende Dissertation ist thematisch und organisatorisch im binationalen Projekt KUBeX (Kollegiales Unterrichtscoaching und Entwicklung experimenteller Kompetenz – Interventionsstudie in der schulpraktischen Ausbildung von Lehrpersonen für Biologie), angesiedelt (cf. Kap. I, 3.1). Sie widmet sich der qualitativen Analyse von Unterrichtsplanungen zum experimentellen Handeln, die angehende Lehrpersonen der Sekundarstufe I

im Rahmen des Projekts KUBeX schriftlich erstellt, in Tandems besprochen und weiterentwickelt haben. Das experimentelle Handeln stellt dabei den eigentlichen Fachinhalt dar. Dieser soll am biologischen Thema «Visuelle Wahrnehmung» erarbeitet werden (Gropengiesser, 2007). Vor dem Hintergrund, dass Beschreibungen der für die Gestaltung eines solchen Experimentalunterrichts notwendigen Expertise derzeit rar sind und entsprechende Messinstrumente fehlen (Hasse et al., 2014), ist im Projekt KUBeX ein neues Erhebungsinstrument entwickelt worden. Dieses basiert auf aktuellen lehr- und lerntheoretischen wie auch fachdidaktischen Erkenntnissen sowie auf dem Verständnis, dass das experimentelle Handeln eine holistische Aktivität darstellt, in der Vor- und Nachbereitung inkludiert sind. Für die Bestimmung der Qualitätsmerkmale, die das fachdidaktische Wissen für die Planung von Unterricht zum experimentellen Handeln operationalisieren, sind deshalb die Wissensfacetten in Anlehnung an das ERTE-Modell normativ festgelegt und eindimensional konzeptualisiert worden. Grundlage dieses Modells bildet das Planungsmodell der «Didaktischen Rekonstruktion». Zur Einlösung des holistischen Ansatzes ist zudem die Reflexion im Sinne einer Ergebnis-, Prozess- und Begriffsreflexion berücksichtigt worden (cf. Börlin, 2012). Planungsentscheidungen und das in diesem Zusammenhang beobachtbare fachdidaktische Wissen zur Planung von Unterricht zum experimentellen Handeln angehender Lehrpersonen sollen mit Hilfe dieses neu entwickelten Strukturmodells rekonstruiert und auch quantifiziert werden. Das Instrument wird in Teilstudie I zunächst allgemein und dann in Bezug auf zwei zentrale Dimensionen des Modells der «Didaktischen Rekonstruktion», «Analyse» und «Konstruktion» (cf. Kap. I, 2.8.4), im Sinne einer explorativen Herangehensweise empirisch geprüft. Die in diesem Zusammenhang konkret ausformulierten Forschungsfragen werden im nächsten Kapitel genauer vorgestellt.

2.11 Forschungsfragen

Die Teilstudie I der vorliegenden Studie verfolgt das Ziel, das im Projekt KUBeX entwickelte Ratinginstrument zur Beurteilung von Unterrichtsplanungen angehender Lehrpersonen in Bezug auf das darin erfassbare fachdidaktische Planungswissen zum experimentellen Handeln hinsichtlich seiner Gütekriterien zu prüfen. Dabei wird das Ratingverfahren auf seine Objektivität untersucht. Ausserdem wird die Reliabilität des Ratingmanuals auf der Grundlage, der im Projekt KUBeX erfassten Daten, ermittelt und das Erhebungsinstrument explorativ validiert. Die Validierung fokussiert in einem ersten Schritt die Eindimensionalität des Konstrukts. Im Fokus dieses Forschungsanliegens 1a steht die Untersuchung, ob das neu entwickelte Instrument tatsächlich das fachdidaktische Wissen zur Gestaltung von Unterricht zum experimentellen Handeln misst. In einem zweiten Schritt wird das Instrument auf die in Kapitel I, 2.8.4 dargelegten Dimensionen, «Analyse» und

«Konstruktion», geprüft (Forschungsanliegen 1b). Dieses Verfahren entspricht einer induktiven, hypothesengenerierenden Vorgehensweise. Für die Teilstudie I ergeben sich vor dem dargestellten Hintergrund daher folgende Forschungsfragen:

1. Lässt sich das fachdidaktische Wissen zur Planung einer Lerngelegenheit bzw. Unterrichtsstunde zum experimentellen Handeln mit Hilfe des im Projekt KUBeX entwickelten hoch inferenten Instruments in Unterrichtsplanungen auf der Grundlage der Daten valide erfassen?
2. Bilden die aus der Beurteilung der Unterrichtsplanungen angehender Lehrpersonen erfassten Daten die zentralen Wissensdimensionen «Analyse» und «Konstruktion» ab, welche im ERTE-Modell enthalten sind?

3 Forschungsdesign und Methoden

Ausgehend von den vorangehend formulierten Fragestellungen werden in diesem Kapitel das Forschungsdesign und die Methoden der Teilstudie I beschrieben. Die vorliegende Studie, die sich aus zwei Teilstudien zusammensetzt (Teilstudie I & II), ist, wie in Kap. 2.10 erwähnt, thematisch und organisatorisch eingebettet in das Projekt KUBeX (Kollegiales Unterrichtscoaching und Entwicklung experimenteller Kompetenz – Interventionsstudie in der schulpraktischen Ausbildung von Lehrpersonen für Biologie). Im Sinne einer Übersicht wird daher zunächst das Projekt KUBeX in seiner Gesamtheit mit den Zielsetzungen sowie dem Projektdesign vorgestellt (Kap. I, 3.1). Anschliessend erfolgt die Präsentation der Teilstudie I der vorliegenden Studie (Kap. I, 3.2). Ein Unterkapitel beleuchtet die Zielsetzungen und das Untersuchungsdesign (Kap. I, 3.2.1). Im Anschluss daran wird die Stichprobe mit Blick auf fachdidaktisch relevante Aspekte für die Analyse der Unterrichtsplanungen zum experimentellen Handeln untersucht (Kap. I, 3.2.2). Das Unterkapitel Durchführung der Untersuchung widmet sich dem Ablauf der Gesamtuntersuchung KUBeX und erläutert das für die Teilstudie I relevante Vorgehen der Videografie der Planungsgespräche (Kap. I, 3.2.3). Im Weiteren wird das Erhebungsinstrument, ein Ratinginstrument beschrieben, mit dessen Hilfe die Performanz und damit das beobachtbare fachdidaktische Wissen zur Planung von Unterricht zum experimentellen Handeln angehender Lehrpersonen in seiner Qualität eingeschätzt werden soll (Kap. I, 3.2.4). Die Analyse der drei zentralen Gütekriterien des Ratinginstruments (Objektivität, Reliabilität und Validität) wird im Anschluss daran detailliert dargestellt. Das Kapitel Design und Methode schliesst mit einer Darlegung der statistischen Auswertungsverfahren der erhobenen Daten in Bezug auf die Forschungsfragen der Teilstudie I.

3.1 Das Projekt KUBeX

Das binationale und interdisziplinäre Projekt KUBeX widmet sich der Untersuchung und Förderung des unterrichtsbezogenen Professionshandelns angehender Lehrpersonen der Sekundarstufe I. Grundlage und Rahmen des Projekts bilden das Modell des Kollegialen Unterrichtscoachings (Kreis & Staub, 2009), eine Form von Peer-Coaching zur Erweiterung der Fähigkeiten und Fertigkeiten von Lehramtsstudierenden in Schulpraktika, sowie das Unterrichtsfach Biologie. Ziel des Projekts ist die Untersuchung und Förderung unterrichtsbezogener Zusammenarbeit der Studierenden bei der Planung und Besprechung von Experimentalunterricht im Fach Biologie mit Hilfe des Content-Focused Peer Coaching-Ansatzes. Die Untersuchung erfolgt mit einem quasi-experimentellen Untersuchungs-

design mit Kontrollgruppenansatz (Abb. 15) und ist in die laufende Ausbildung angehender Sekundarlehrpersonen eingebettet. Es sind insgesamt 119 Studierende (Lehrpersonenausbildung Biologie Sekundarstufe I) an der Untersuchung beteiligt, davon 66 Studierende an der Pädagogischen Hochschule Weingarten (Deutschland) und insgesamt 53 Studierende der Pädagogischen Hochschulen Thurgau, St. Gallen und Zürich (Schweiz; $n_{\text{Thurgau}} = 8$; $n_{\text{St. Gallen}} = 23$; $n_{\text{Zürich}} = 22$) (cf. Tab. 4). Die 119 Studierenden sind mehrheitlich randomisiert je einer Kontroll- bzw. Interventionsgruppe zugeteilt worden. Erstere umfasst 54, letztere 65 Studierende. Die unterschiedliche Gruppengrösse ist u.a. darauf zurückzuführen, dass die wenigen Studierenden der Pädagogischen Hochschule Thurgau aus organisatorischen Gründen der Interventionsgruppe zugeteilt worden sind.

Das Kernstück der Untersuchung bildet eine Modulveranstaltung (Intervention) zum Thema «Kollegiales Unterrichtscoaching» (Kreis & Staub, 2009) mit Fokus auf dem Content-Focused Peer Coaching, deren Wirkung auf das Planungshandeln von Studierenden der Sekundarstufe I geprüft werden soll. Nur die Interventionsgruppe nimmt an dieser Veranstaltung teil, die Kontrollgruppe erhält währenddessen einen fachlich unabhängigen Input, der thematisch nicht interferiert. Alle Studierenden sowohl der Interventions- wie auch der Kontrollgruppe sind angehalten, nach der Modulphase eine Unterrichtsplanung auf der Grundlage einer Vignette zu erstellen (cf. Anhang A-1, Abb. A1.1). Den Fachinhalt für diese Unterrichtsplanung bildet das experimentelle Handeln im Sinne der Erkenntnisgewinnung, das am biologischen Inhalt «Visuelle Wahrnehmung» (Gropengiesser, 2007) erarbeitet werden soll. Das experimentelle Handeln, d.h. das Experimentieren im Sinne der Erkenntnisgewinnung, ist in den Bildungsstandards verankert (EDK, 2011; KMK, 2005) und damit verbindlich für den aktuellen Naturwissenschaftsunterricht sowohl in Deutschland wie auch in der Schweiz. Das biologische Thema «Visuelle Wahrnehmung» ist in beiden Ländern Teil der Curricula der Sekundarstufe I. Es kann unabhängig von saisonalen Bedingungen unterrichtet werden und gewährleistet damit die Vergleichbarkeit über die Grenzen. Eine fachdidaktische Impulsveranstaltung zum experimentellen Handeln, die vor der Intervention von allen 119 Studierenden besucht werden muss, soll das Experimentieren im Sinne der Erkenntnisgewinnung nochmals thematisieren und damit die Vergleichbarkeit der Pädagogischen Hochschulen erhöhen.

Die von allen Studierenden aufgrund einer Vignette (cf. Anhang A-1, Abb. A1.1) erstellten Planungen werden im Rahmen eines videografierten Planungsgesprächs in Peergroups gegenseitig vorgestellt, diskutiert und weiter elaboriert. Zur Untersuchung möglicher Wechselwirkungen zwischen den erstellten Unterrichtsplanungen und der professionellen Handlungskompetenz wird ausserdem das Professionswissen der Studierenden an drei Messzeitpunkten online erhoben.

Unterschiede zwischen den Experimentalgruppen werden dabei mit Bezug auf die Intervention hinsichtlich des Fachwissens, fachdidaktischen Wissens und der Überzeugungen quantitativ und qualitativ untersucht.

Die Überprüfung der Wirkung der Intervention zum Content-Focused Peer Coaching auf das Planungshandeln der Studierenden erfolgt mit einem multi-methodischen Ansatz. Zu den im Projekt KUBeX eingesetzten Erhebungsinstrumenten gehören Fragebogenerhebungen und Tests, schriftlich verfasste Unterrichtsplanungen, videografierte dyadische Planungsgespräche, Dokumente zu den Kontextvariablen sowie Expertenbefragungen (cf. Abb. 15). Mit Hilfe der Fragebogenerhebungen und Tests wird die professionelle Handlungskompetenz der Studierenden in seinen Facetten erhoben. Das Planungshandeln wird auf der Grundlage der videografierten, dyadischen Unterrichtsplanungsgespräche der Studierenden und der schriftlich erstellten Unterrichtsplanungen und -materialien untersucht. Kontextvariablen werden mittels Dokumentenanalyse ermittelt, eine Expertenbefragung soll ausserdem Einblick in die Ausbildungen an den vier verschiedenen Pädagogischen Hochschulen liefern. Eine genaue Beschreibung der Durchführung des Gesamtprojekts erfolgt im Rahmen der Ausführungen zur Teilstudie I der vorliegenden Studie in Kapitel I, 3.2.3. Teilstudie I der vorliegenden Studie ist Gegenstand des nächsten Kapitels.

3.2 Teilstudie I – Validierung eines Ratinginstruments zur Analyse von Unterrichtsplanungen, -materialien und videografierten Planungsgesprächen

Während im Gesamtprojekt KUBeX das Planungshandeln mittels einer Intervention gefördert und untersucht wird, widmet sich die vorliegende, fachdidaktisch ausgerichtete qualitative Studie (Teilstudie I & II) den im Projekt entstandenen Unterrichtsplanungen, -materialien und videografierten Planungsgesprächen, die, wie in Kapitel I, 2.8.2 dargelegt, einen möglichen Zugang zur Erfassung des fachdidaktischen Wissens von angehenden Lehrpersonen darstellen (Baer et al., 2011; Stender, Brückmann & Neumann, 2017; Terbrügge, 2001; Tillema, 2009). Im Zentrum der Teilstudie I steht die Validierung des im Projekt KUBeX entwickelten Ratingmanuals. Die Validierung erfolgt auf der Basis der Daten, die aufgrund der Analyse der Performanz experimentellen Handelns in Unterrichtsplanungen und -materialien sowie durch die direkte Beobachtung der videografierten Gespräche über die Unterrichtsplanungen generiert worden sind. Die Teilstudie I wird in den nächstfolgenden Abschnitten vorgestellt. Damit verbunden wird auch die Durchführung des Gesamtprojekts KUBeX beschrieben. Dabei wird ein besonderes Augenmerk auf die für die Teilstudie I wichtige Datenerhebung und -auswertung gelegt.

3.2.1 Zielsetzungen und Design

Für die Entwicklung fachdidaktischen Wissens, das als wichtige Grundlage für die Ausübung der Lehrerprofession angesehen wird (Kind, 2009), kommt dem Planungshandeln eine wesentliche Bedeutung zu (Ball & Bass, 2000; Baylor, 2002). Beschreibungen der dafür erforderlichen Expertise, insbesondere zum Planen von Unterricht zum experimentellen Handeln als Fachinhalt sind derzeit jedoch rar. Entsprechend sind geeignete Erhebungsinstrumente, mit deren Hilfe eine standardisierte Erfassung der Qualität der Planungsperformanz in Bezug auf das beobachtbare fachdidaktische Wissen erfolgt, selten und inkonsistent (Hasse et al., 2014; Vogelsang & Riese, 2017; Weingarten & van Ackeren, 2017). Vor diesem Hintergrund ist im Projekt KUBeX ein Ratingmanual entwickelt worden, mit dessen Hilfe die Unterrichtsplanungen und -materialien, aber auch die videografierten Gespräche über die Unterrichtsplanungen analysiert werden sollen. Ziel der Teilstudie I der vorliegenden Studie ist zu prüfen, ob das theoriebasiert entwickelte Erhebungsinstrument das fachdidaktische Wissen zur Planung von Unterricht zum experimentellen Handeln als eindimensionales Konstrukt valide erfassen kann und ob sich die beiden Wissensdimensionen «Analyse» und «Konstruktion», die im Modell der «Didaktischen Rekonstruktion» enthalten sind (Kap. I, 2.8.4), in den erfassten Daten abbilden.

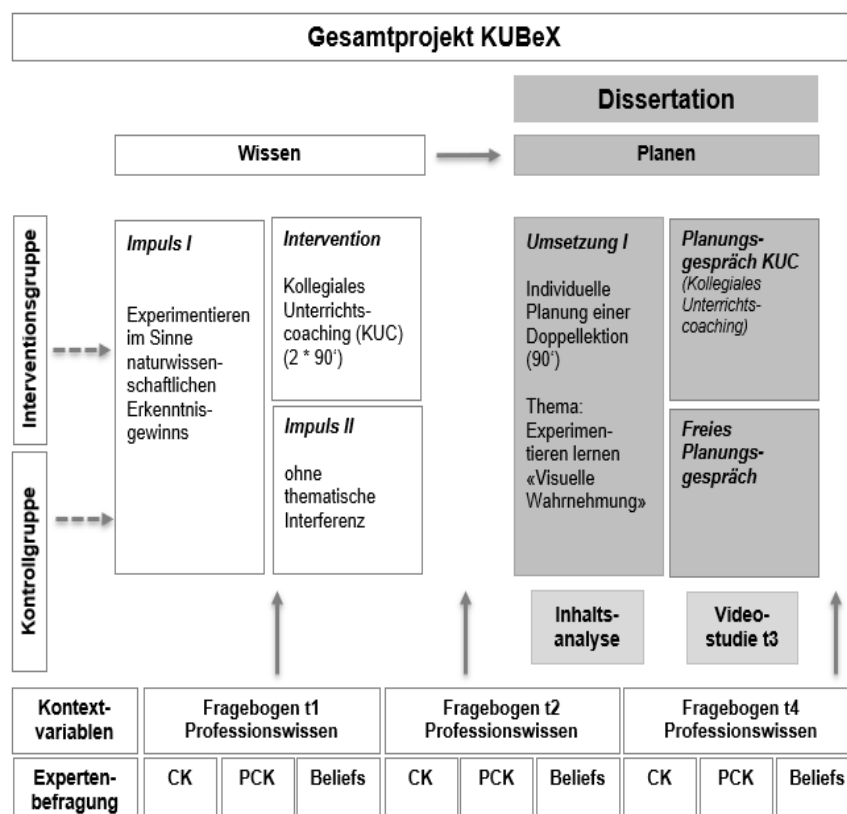


Abbildung 15: Design der Teilstudie I eingebettet in das Gesamtdesign des Projekts KUBeX, Teilstudie I (grau hinterlegt): Datenquellen (dunkelgrau), Erhebungsverfahren (hellgrau)

Das auf diese Zielsetzungen ausgerichtete Design der Teilstudie I ist in Abbildung 15, eingebettet in das Gesamtprojekt KUBeX, dargestellt. Auch sind in der Grafik die für die Teilstudie I wesentlichen Datenquellen und Erhebungsverfahren in zwei verschiedenen Grautönen speziell hervorgehoben. Die Datenquellen und Erhebungsverfahren werden in Kap. I, 3.2.3 zusammen mit der Durchführung des Gesamtprojekts KUBeX näher erläutert. Vorangehend wird zunächst aber noch die Stichprobe des Gesamtprojekts KUBeX und damit auch beider Teilstudien (Teilstudie I & II) der vorliegenden Untersuchung genauer beschrieben.

3.2.2 Stichprobe

Die Stichprobe des Gesamtprojekts KUBeX, an dem insgesamt 119 Studierende von vier Pädagogischen Hochschulen beteiligt waren (Kapitel I, 3.1), bildete die Grundlage für Teilstudie I, sowie auch für Teilstudie II. Etwas mehr als die Hälfte dieser Studierenden absolvierte die Ausbildung in Weingarten (Tab. 4).

Das Durchschnittsalter der Studierenden betrug 22.9 Jahre ($SD = 4.3$), zwei Drittel von ihnen waren weiblich, ein Drittel männlich.

Tabelle 4: Detaillierte Stichprobenbeschreibung der vorliegenden Studie (Teilstudie I & II)

		absolute Häufigkeit	relative Häufigkeit (%)	<i>MW</i>	<i>SD</i>
Pädagogische Hochschule					
	Thurgau	8	6.7		
	St. Gallen	23	19.3		
	Zürich	22	18.5		
	Weingarten	66	55.5		
Experimentalgruppe					
	Intervention	65			
	Kontroll	54			
Geschlecht	m	30	25.2		
	w	89	74.8		
Alter				22.9	3.4
Stand Ausbildung (Semesterzahl)				5.2	1.5
	bereits Biologie unterrichtet	91	76.5		
	bereits im Unterricht experimentiert	64	53.8		

Im Schnitt haben alle Studierenden 5.2 Semester ($SD = 1.5$) studiert. 76.5% dieser Studierenden haben bereits Biologie unterrichtet, 53.8 % sogar in ihrem Unterricht experimentiert.

3.2.3 Durchführung

Die Durchführung des Gesamtprojekts KUBeX und der darin eingebetteten vorliegenden Studie (Teilstudie I & II) ist im Frühlingsemester 2014 erfolgt. Sie dauerte approximativ einen Monat und wurde an jeder der vier Pädagogischen Hochschulen durch dasselbe Forcherteam nach einem standardisierten Verfahren realisiert. Gestartet wurde mit einem ersten von drei Professionswissenstests, den alle am Projekt beteiligten 119 Studierenden im Rahmen einer regulären fachdidaktischen Modulveranstaltung online am eigenen PC lösen mussten. In der darauffolgenden Woche fand die Impulsveranstaltung I statt, in der sich die Studierenden 2×90 Minuten mit fachlichen und fachdidaktischen Aspekten in Bezug auf das experimentelle Handeln und damit in Bezug auf den Prozess der Erkenntnisgewinnung auseinandergesetzt haben (ergänzende Informationen cf. Smit, Weitzel, et al., 2017). Diese Impulsveranstaltung I war ökologisch annähernd valide, weil sie im Rahmen einer Modulveranstaltung der Studierenden stattgefunden hat (Smit, Rietz & Kreis, 2017). Daran anschliessend fand ein zweiter Online-Test statt, der innerhalb einer Woche am eigenen PC bearbeitet werden musste. Studierende wurden in der Folge randomisiert der Interventions- bzw. Kontrollgruppe zugeteilt. Es wurden in beiden Gruppen Tandems im Sinne von kleinen Peer-Groups gebildet. Die Interventionsgruppe erhielt einen Input zum Content-Focused Peer Coaching (2×90 Minuten), die Kontrollgruppe zu einem zusätzlichen fachdidaktischen Impuls II (2×90 Minuten), der nicht mit dem Fokus des Projekts interferierte. Sowohl der fachdidaktische Impuls II als auch die Intervention wurden an allen vier Pädagogischen Hochschulen durch die gleichen Fachpersonen erteilt. Für beide Inputveranstaltungen lag je ein im Projekt entwickeltes und verbindliches Skript vor. Aufgrund einer vorgegebenen Vignette (cf. Anhang A-1, Abb. A1.1), in der wichtige Rahmenbedingungen zum Schulhaus, zur Klasse, aber auch für die Planungsaufgabe festgehalten waren, mussten die Studierenden eine Doppellektion zum Thema «Experimentieren-lernen am Themenbereich «Visuelle Wahrnehmung» (Gropengiesser, 2007)» planen. Die Themen Adaptation oder Akkommodation standen zur Auswahl. Jeder Studierende eines Tandems wählte je ein Thema aus und entwickelte dazu in max. drei Arbeitsstunden eine Unterrichtsplanung für eine Doppellektion in digitaler Form. Die Planung musste anhand einer digitalen Planungsvorlage erfolgen, die zwar eine gewisse Struktur vorgab, aber dennoch nicht alle, bei der Planung zu berücksichtigenden Aspekte (cf. Anhang A-1, Abb. A1.2). Mögliche Ideen, Materialien, welche die Studierenden in den Unterricht einbeziehen wollten, sollten sie

kopiert in die Planungsgespräche mitbringen. Die Planungsgespräche fanden in der darauffolgenden Woche gemäss eines vorher festgelegten Plans statt. Für die Planungsgespräche wurde an jeder Pädagogischen Hochschule ein speziell dafür geeigneter Raum mit einem Tisch und zwei Kameras mit Unterstützung der Mediendienste eingerichtet. In Anlehnung an Erfahrungen aus der TIMSS-Videostudie 1999 (Jacobs et al., 2003) und der IPN-Videostudie Physik (Seidel, Prenzel, Duit & Lehrke, 2003), aber auch von Kreis und Staub (2007) wurden zwei Videokameras eingesetzt, eine fix installierte Front- und eine ebenso fest installierte Deckenkamera, beide ausgestattet mit guten Mikrofonen (Abb. 16).

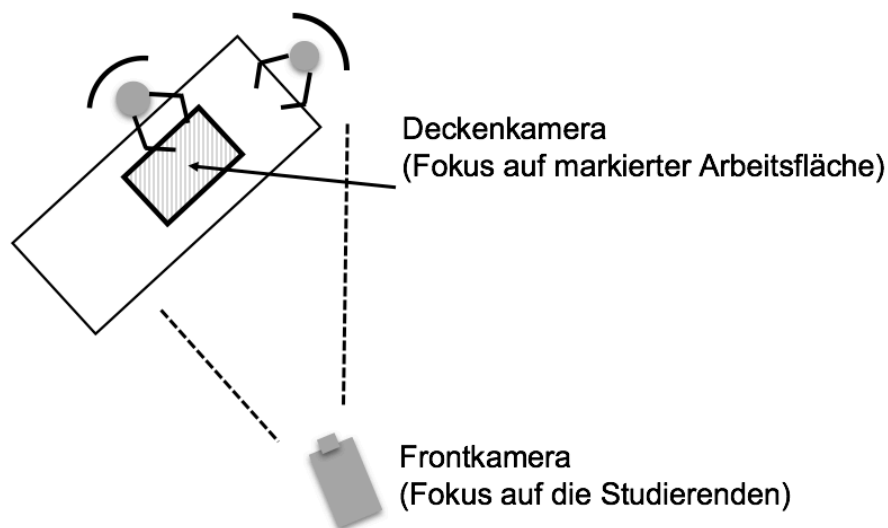


Abbildung 16: Kamerapositionen für die Videoaufzeichnung der Planungsgespräche

Die Frontkamera war auf beide Studierenden gerichtet, um den gesamten Gesprächsverlauf des Planungsgesprächs dokumentieren zu können. Die Deckenkamera wurde so installiert, dass sie die auf dem Tisch mit Klebeband markierte Fläche fokussierte, in welche die Studierenden ihre digital erstellten Planungen oder aber auch Materialien während des Gespräches legen sollten, und somit alle Ereignisse im Zusammenhang mit diesen Materialien festhielt. Den Studierenden standen für die Planungsgespräche jeweils max. 45 Minuten pro Thema zur Verfügung. In dieser Zeit waren sie entweder selbst Coach oder wurden bei der Besprechung der eigenen Unterrichtsplanung durch den Tandempartner oder die -partnerin gecoacht und waren Coachee. Der Ablauf der videografierten Planungsgespräche gestaltete sich so, dass zuerst der Coachee seinen Code angeben und mit einem Zeichen (einem Händeklatsch) den Start des Gesprächs markieren musste. Mit Hilfe dieses

akustischen Zeichens konnten die Aufnahmen der Decken- und der Frontkamera zu einem späteren Zeitpunkt zeitlich synchronisiert werden. Dann stellte der Coachee seine Unterrichtsplanung vor, bevor diese mit dem Coach diskutiert und teilweise auch elaboriert wurde. Die in der Diskussion entstandenen Ideen, Anpassungen oder Ergänzungen mussten handschriftlich auf der digital erstellten Unterrichtsplanung ergänzt werden. So konnten Veränderungen im Sinne von Anpassungen oder Elaborationen von der ursprünglichen Planung unterschieden werden. Nach erfolgreichem erstem Planungsgespräch wurden die Rollen getauscht und der Coach übernahm die Rolle des Coachees und präsentierte in der gleichen Weise seine Planung. Im Anschluss an die Coachinggespräche fand ein dritter Onlinetest statt, der von den Studierenden wiederum am PC innerhalb einer Woche gelöst werden musste.

Die Datengrundlage für Teilstudie I bilden die Unterrichtsplanungen und -materialien sowie auch die videografierten, in Tandem oder Peers durchgeführten Planungsgespräche zu den erstellten Unterrichtsplanungen. Dabei soll das im Projekt KUBeX entwickelte Erhebungsinstrument zur Einschätzung der Performanz und damit der Qualität fachdidaktischen Wissens zur Gestaltung von Unterricht zum experimentellen Handeln als Fachinhalt im Hinblick auf die Gütekriterien geprüft werden. Das Instrument (Ratingmanual) wird daher im nächsten Abschnitt genauer vorgestellt.

3.2.4 Ratingmanual zur Erfassung des fachdidaktischen Wissens zur Planung von Unterricht zum experimentellen Handeln

Im Rahmen des Projekts KUBeX ist ein Ratingmanual entwickelt worden, mit dessen Hilfe die Qualität des fachdidaktischen Wissens in Unterrichtsplanungen erfasst und in gewisser Weise die Planungskompetenz angehender Lehrpersonen hinsichtlich des experimentellen Handelns eingeschätzt werden kann. Die Entwicklung ist nach dem qualitativen und inhaltsanalytischen Verfahren der Strukturierung nach Mayring (2015) erfolgt. Ziel einer solchen Analyse ist es, eine bestimmte Struktur mit Hilfe eines Kategoriensystems aus dem Material herauszufiltern. Dabei werden vier Formen der Strukturierung, die formale, die inhaltliche, die typisierende und die skalierende Strukturierung unterschieden (Mayring, 2015, S. 99). In der vorliegenden Studie kam die skalierende Strukturierung zur Anwendung, bei der es darum geht, das Material bzw. bestimmte Materialanteile auf einer Skala (in der Regel Ordinalskala) einzuschätzen (Mayring, 2015, S. 106). Ausgehend von empirischer Literatur wurde ein Kategoriensystem im Sinne eines Kompetenzstrukturmodells entwickelt (cf. Kap. 2.4.2), mit dem sich das fachdidaktische Wissen zur Planung von Unterricht zum experimentellen Handeln operationalisieren lässt und mit dessen Hilfe die Unterrichtsplanungen, -materialien, aber auch die videografierten Planungsgespräche ana-

lysiert werden können. Die Entwicklung des Ratingmanuals wird im nächsten Abschnitt beschrieben.

I. Entwicklung des Ratingmanuals

Grundlage für die Entwicklung dieses Kompetenzstrukturmodells und seiner Kategorien bildete das ERTE-Modell (Kizil & Kattmann, 2013; Kap. I, 2.8.5). Dieses beruht auf dem für die Fachdidaktik Biologie zentralen Planungs- und Forschungsmodell der «Didaktischen Rekonstruktion» (van Dijk & Kattmann, 2007, 2010; Duit, Gropengiesser, Kattmann, Komorek & Parchmann, 2012; Kattmann, Duit, Gropengiesser & Komorek, 1997; Kap. I, 2.8.4) und beschreibt das fachdidaktische Wissen, das in Beziehung gesetzt wird zu den Kompetenzansprüchen der Fachdidaktik hinsichtlich fachlicher Klärung, Schülervorstellungen und didaktischer Strukturierung der Lerngelegenheit (Kizil & Kattmann, 2013; Kap. I, 2.8.5). Die Auswahl der Kategorien für das Ratingmanual erfolgte daher auch nach Ansätzen konstruktivistischen Lehr- und Lernverständnisses. Diese scheinen gemäss Widodo und Duit (2004, S. 233) insgesamt gesehen besser geeignet zu sein, Schülerinnen und Schüler zu einer naturwissenschaftlichen Grundbildung im Sinne von *Scientific Literacy* zu führen. Für die Entwicklung des Kategoriensystems wurden ausserdem die folgenden Aspekte berücksichtigt:

- die Bildungsstandards (KMK, 2005; EDK, 2011) und die sich daraus für die Lehrpersonenbildung ergebenden Aufgaben
- bereits bestehende Ratingsysteme zur Analyse von Unterricht zum experimentellen Handeln oder zur Planung eines solchen Unterrichts (Börlin, 2012; Hasse et al., 2014)
- aktuelle Ergebnisse zu naturwissenschaftlichen Denk- und Arbeitsweisen, insbesondere zur Kompetenzstruktur und Kompetenzentwicklung naturwissenschaftlicher Erkenntnisgewinnung (e.g. Mayer, 2007; Mayer, Grube & Möller, 2008; Priemer, 2011; Schreiber, 2012; Wellnitz, 2012)
- Erkenntnisse zu den Phasen experimentellen Handelns und damit zu den Teilprozessen der Erkenntnisgewinnung (Hammann, Phan & Bayrhuber, 2008; Mayer et al., 2008; Nowak, Nehring, Tiemann & Belzen, 2013; Wellnitz et al., 2017; Wellnitz & Mayer, 2013)
- Kenntnisse von Schwierigkeiten, die angehende oder erfahrene Lehrpersonen bei der Planung und Durchführung von Experimentalunterricht bekunden (e.g. Capps & Crawford, 2013b; Chabalengula, Mumba & Mbewe, 2012; Davis, Petish & Smithy, 2006; Fazio et al., 2010; Gyllenpalm & Wickman, 2011b; Kunz, 2011; Roehrig & Luft, 2004; Schneider & Plasman, 2011; Wallace & Kang, 2004)
- Empirische Befunde zu möglichen Fehlern, die von Schülerinnen und Schülern beim Experimentieren gemacht werden (e.g. Hammann, 2004; Hasse, Joachim, Bögeholz & Hammann, 2014).

Wie bereits in Kap. I, 2.7.2 erläutert, liegt der Fokus des experimentellen Handelns auf der wissenschaftstheoretischen Dimension, genau genommen auf dem Teilprozessansatz experimenteller Kompetenz. Das experimentelle Handeln wird in dieser Studie als holistische Aktivität (Börlin, 2012; Hodson, 2014; Millar, 2004) verstanden. Es inkludiert neben dem Experimentieren auch die Vor- und Nachbereitung und wird deshalb in Anlehnung an Börlin als Experimentiereinheit definiert (Börlin, 2012, S. 51). Eine solche Einheit kann in den Unterrichtsplanungen je nach der im Unterricht fokussierten Zielsetzung eine ganze Doppelktion in Anspruch nehmen oder aber nur Teilphasen des Unterrichts betreffen. Auch kann sie den gesamten Experimentalprozess beinhalten oder aber Teilprozesse des experimentellen Handelns fokussieren, wie z.B. die Formulierung von überprüfbaren Fragestellungen. Das fachdidaktische Wissen zur Planung von Unterricht zum experimentellen Handeln wird mit Hilfe des Kategoriensystems jeweils dahingehend eingeschätzt, ob und in welcher Qualität die Experimentiereinheit einen Beitrag zur Erarbeitung relevanter Konzepte des Erkenntnisprozesses mit Schülerinnen und Schülern leistet. Die ausgewählten, fachlich und fachdidaktisch begründeten Kriterien umschreiben eine Planung zum experimentellen Handeln von hoher Qualität. Die Kriterien basieren auf theoretischen und empirischen Erkenntnissen (cf. Kap. I, 3.2.4). Da es sich beim experimentellen Handeln um ein latentes Merkmal handelt, das nicht direkt beobachtbar ist, steht die Interpretation des Beobachters bei diesem Vorgehen im Vordergrund. Bisherige Videostudien lassen aber vermuten, dass sich sogenannte hoch inferente Methoden wie diese, dazu eignen, qualitätsrelevante Aspekte des Unterrichts zu entschlüsseln (cf. Börlin, 2012, S. 56).

Das Ratingmanual für die hoch inferente Analyse der Unterrichtsplanungen und videografierten Planungsgespräche wurde in einem iterativen Verfahren entwickelt (Abb. 17). In einem ersten Schritt wurde die aktuelle Literatur geprüft und analysiert (deduktiv). Ausgehend von den Erkenntnissen aus der Literatur wurde in einem zweiten Schritt eine Rohfassung eines kategoriengeleiteten Ratingsystems erstellt, mit dessen Hilfe sich wesentliche Aspekte experimentellen Handelns in Planungen und im Unterricht ermitteln lassen. Diese Rohfassung des hoch inferenten Ratingmanuals wurde mit Fachkolleginnen und Fachkollegen diskutiert und gemäss der Kritikpunkte angepasst. Dabei wurden Indikatoren ergänzt oder gestrichen. Ausserdem wurden die a priori definierten und ausformulierten Niveau- bzw. Qualitätsbeschreibungen der Indikatoren sprachlich ausgeschärft. Mit Hilfe von Beispielen videografierten Planungsgespräche wurde das Manual in einem nächsten Schritt von zwei unabhängigen Ratern ein weiteres Mal kritisch geprüft (induktiv). Unklarheiten wurden behoben, ungenaue oder missverständliche Formulierungen korrigiert.

Mit Hilfe dieses qualitativen Verfahrens ist im Projekt KUBeX ein Manual zur Erfassung des fachdidaktischen Wissens zur Planung von Unterricht zum experimentellen Handeln entwickelt worden (Weitzel, 2015). Der Aufbau dieser in mehrfachem Wechsel zwischen induktiver und deduktiver Herangehensweise entstandenen Arbeitsversion des Ratingmanuals wird im nächsten Abschnitt vorgestellt.

II. Aufbau des Ratingmanuals

Das kategoriengeleitete Ratingmanual zur Erfassung des fachdidaktischen Wissens angehender Lehrpersonen zur Planung von Unterricht zum experimentellen Handeln umschreibt eine idealtypische Unterrichtsplanung und damit im Sinne des Angebots-Nutzungs-Modells von Helmke (2012; Kap. I, 2.6.1) die Qualität eines Angebots (Abb. 19).

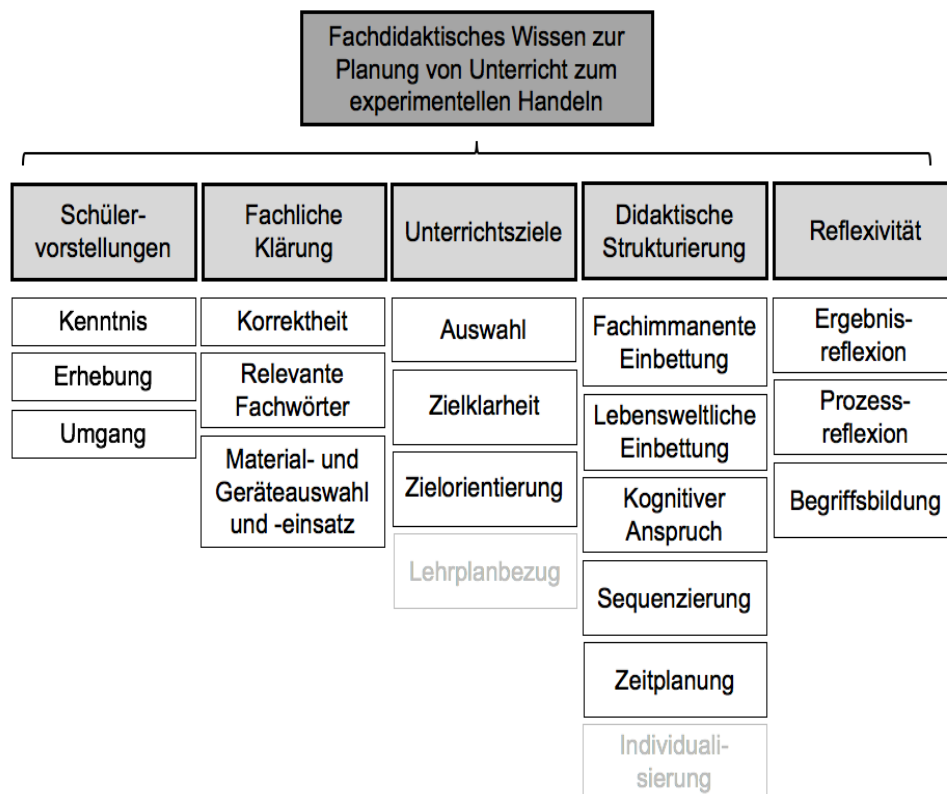


Abbildung 19: Ratingmanual zur Analyse des fachdidaktischen Wissens zur Planung von Unterricht zum experimentellen Handeln (Das Manual wurde im Projekt KUBeX entwickelt, Weitzel, 2015; grau: Indikatoren, die im psychometrischen Messmodell nicht berücksichtigt worden sind)

Das Manual beinhaltet, abgeleitet vom ERTE-Modell von Kizil und Kattmann (2013; Kap. I, 2.8.5), das auf dem Modell der «Didaktischen Rekonstruktion» basiert (van Dijk & Kattmann, 2007; Duit, Gropengiesser, Kattmann, Komorek & Parchmann, 2012; Kattmann, Duit, Gropengiesser & Komorek, 1997), vier zentrale Kategorien fachdidaktischen Wissens. Dazu gehört das Wissen zu den «Schülervorstellungen», zum «Fachwissen» im

Sinne einer fachlichen Klärung, zu den «Unterrichtszielen» sowie zur «Didaktischen Strukturierung» des geplanten Unterrichts. Die ersten drei Kategorien verkörpern die Wissensdimension der «*Analyse*». Die didaktische Strukturierung hingegen entspricht der Dimension der «*Konstruktion*». Es handelt sich hierbei um die im Modell der «Didaktischen Rekonstruktion» beschriebene Sachstruktur des Unterrichts, die sich aus der Interdependenz von fachlicher Klärung, Schülervorstellungen und auch Zielvorgaben ergibt. Das Manual wurde ergänzt durch eine fünfte Kategorie, die «Reflexivität», in Anlehnung an Börlin (2012). Dieser hat in seiner Arbeit – ausgehend von Hodson (1996), der das experimentelle Handeln als reflexive Aktivität erachtet – die Reflexivität als eigene Perspektive postuliert, weil erst mit der Reflexion eine Entwicklung für die Schülerinnen und Schüler sichtbar wird.

Die insgesamt fünf Kategorien wurden durch Indikatoren im Sinne von Qualitätsaspekten konkretisiert, die bei der Planung von Unterricht innerhalb dieser Kategorie berücksichtigt werden sollten. Dabei sind die fünf Kategorien jeweils mit drei bis sieben Indikatoren, insgesamt 19 Indikatoren, ausdifferenziert worden (Abb. 19). Zur Ausdifferenzierung wurden bestehende Instrumente zur Einschätzung der Qualität dieser Kategorien beigezogen und adaptiert. Des Weiteren wurden auch relevante theoretische Grundlagen sowie empirische Erkenntnisse dazu berücksichtigt. Die Literaturbezüge zu jedem Indikator sind in Tabelle 5 zusammengestellt. Die Tabelle erhebt keinen Anspruch auf Vollständigkeit, sondern soll die Auswahl der Indikatoren mit einigen zentralen Quellen belegen.

Tabelle 5: Indikatoren des Ratingmanuals mit ausgewählten Literaturbezügen

Kategorien	Indikatoren	Literatur
Schülvorstellungen	Kenntnis	Duit, Gropengiesser, Kattmann, Komorek & Parchmann (2012); Kattmann et al. (1997); Rakoczy & Pauli (2006); Stender (2014)
	Erhebung	Adamina, Wyssen, Möller, Steffensky & Sunder (2017); Duit et al. (2012); Goldston et al. (2013); Jacobs, Martin & Otieno (2008); Kattmann et al. (1997); Rakoczy & Pauli (2006); Vogelsang (2014); Widodo & Duit (2004)
	Umgang	Duit et al. (2012); Gardner & Gess-Newsome (2011); Jacobs et al. (2008); Kattmann et al. (1997); Neuhaus (2007); Rakoczy & Pauli (2006); Schulz (2011); Vogelsang (2014); Widodo & Duit (2004)
Fachliche Klärung	Korrektheit	Hammann, (2004); Hammann, Phan & Bayrhuber (2008); Kempa (1986); Mayer (2007); Mayer, Grube & Möller (2008); Riese (2009); Schreiber (2012)
	Relevante Fachwörter	Adamina et al. (2017); Neuhaus (2007); Wüsten (2010)
	Material- und Geräteauswahl und -einsatz	Goldston et al. (2013); Kempa (1986); Ruys, Keer & Aelterman (2012); Schreiber (2012)
Ziele	Auswahl der Ziele	Goldston et al. (2013); Jacobs et al. (2008); Schreiber (2012); Stender (2014)
	Zielorientierung	Herweg (2008); Jacobs et al. (2008); Prenzel et al. (2002); Seidel et al. (2006); Stender (2014)
	Zielklarheit	Adamina et al. (2017); Börlin (2012); Rakoczy & Pauli (2006); Seidel, Rimmele & Prenzel (2005); Vogelsang (2014)
	Lehrplanbezug	EDK (2011); Jacobs et al. (2008); KMK (2005); Schulz (2011)
Didaktische Strukturierung	Fachimmanente Einbettung	Börlin (2012); Riese (2009); Stender (2014); Tesch (2005); Tesch & Duit (2004); Vogelsang (2014)
	Lebensweltliche Einbettung	Börlin (2012); Habig, van Vorst & Sumfleth (2018); Jacobs et al. (2008); Walpuski & Schulz (2011); Wüsten (2010)
	Kognitiver Anspruch	Adamina et al. (2017); Baumert & Kunter (2011b); Börlin (2012); Jatzwauk, Rumann & Sandmann (2008); Priemer (2011); Rakoczy & Pauli (2006); Stender (2014); Vogelsang (2014); Widodo & Duit (2004)
	Zeitplanung	Carnahan (1980); Goldston et al. (2013); Riese (2009); Ruys et al. (2012)
	Sequenzierung	Adamina et al. (2017); Hardy, Jonen, Möller & Stern (2006); Ohle, Fischer & Kauertz (2011); Seidel et al. (2005); Vogelsang (2014); Walpuski (2006); Wüsten (2010)
	Individualisierung	Jacobs et al. (2008); Goldston et al. (2013); Rakoczy & Pauli (2006); Riese (2009); Walpuski & Schulz (2011)
Reflexivität	Ergebnisreflexion	Börlin (2012); Bodzin & Beerer (2003); Bybee (2002); Gardner & Gess-Newsome (2011); Jacobs et al. (2008); Kempa (1986); Klahr (2000); Lunetta, Hofstein & Clough (2005); Ruys et al. (2012); Walpuski & Schulz (2011)
	Prozessreflexion	Börlin (2012); Bodzin & Beerer (2003); Bybee (2002); Gardner & Gess-Newsome (2011); Höttecke & Riess (2015); Kempa (1986); Klahr (2000); Lunetta et al. (2005); Riese (2009); Ruys et al. (2012)
	Begriffsbildung	Börlin (2012); Bybee (2002)

Das Kompetenzstrukturmodell und Ratinginstrument besteht zusammengefasst aus 5 Kategorien und insgesamt 19 Kompetenzfacetten bzw. Indikatoren, welche die korrespondierenden fachdidaktischen Wissensfacetten, um einen Unterricht zum experimentellen Handeln zu planen, in ihrer Ausprägung qualitativ festlegen. Die Indikatoren wurden jeweils mit Hilfe einer vierstufigen Rating-Skala von *trifft nicht zu* (1 Punkt) bis *trifft völlig zu* (4 Punkte) hinsichtlich ihrer Qualität graduell abgestuft beschrieben. Das bedeutet, dass die Übereinstimmung mit dem Idealtyp eines Merkmals oder Indikators umso stärker ist, je höher der numerische Ausdruck ist (cf. Rakoczy & Pauli, 2006, S. 209). Die Äquidistanz zwischen den verschiedenen Qualitätsstufen wird in Anlehnung an entsprechende Untersuchungen angenommen (e.g. Börlin, 2012), obgleich es sich im Grunde um ordinale Zahlen handelt. Aus der Einschätzung aller Indikatoren (manifeste Variablen) und damit einem erhobenen Qualitätswert, sollte sich die Qualität des fachdidaktischen Wissens für die Planung von Unterricht zum experimentellen Handeln im Sinne einer möglichst umfassenden Beschreibung der erforderlichen Kompetenz (latente Variable) erfassen lassen.

Im Laufe der Entwicklung des kategoriengeleiteten Ratingmanuals hat sich allerdings gezeigt, dass zwei Indikatoren, nämlich der «Lehrplanbezug» und die «Individualisierung» in keiner der untersuchten Unterrichtsplanungen bzw. keinem der Planungsgespräche aufgetreten sind. Sie wurden deshalb im psychometrischen Messmodell nicht mehr berücksichtigt (Abb. 19).

III. Datenaufbereitung und -analyse mit Hilfe des Ratingmanuals

Für die Analyse der Unterrichtsplanungen, -materialien sowie der videografierten Planungsgespräche mussten die Daten entsprechend aufbereitet werden. Alle von den Studierenden digital erstellten Unterrichtsplanungen sowie Arbeitsmaterialien wurden kopiert und eingescannt. Die mit Hilfe der Codes aus Datenschutzgründen anonymisierten Videoaufnahmen der Front- und Deckenkamera wurden synchronisiert und die daraus zu entnehmenden Gespräche transkribiert. Die Einschätzung, inwiefern die Unterrichtsplanungen, -materialien sowie videografierten Planungsgespräche mit einer idealtypischen Planung übereinstimmen, erfolgte durch zwei unabhängige Kodierer und Rater, um das Ausmass der Subjektivität bei den Ratings zu kontrollieren (cf. Brückmann & Duit, 2014). Beide Rater haben an der Entwicklung des Manuals mitgearbeitet und sind ausserdem in der fachdidaktischen Ausbildung angehender Sekundarlehrpersonen in Natur und Technik tätig. Aufgrund dieser Voraussetzungen waren sie mit dem Manual vertraut. Die Kodierung der Unterrichtsplanungen, -materialien und videografierten Planungsgespräche wurde eventbasiert durchgeführt, um die Experimentiereinheiten als Sinneinheiten erfassen zu können und Verzerrungen zu vermeiden (cf. Brückmann & Duit, 2014). Sie verlief immer in der

gleichen Abfolge. In einem ersten Schritt wurde die in der Planung vorzufindene Experimentiereinheit, die einen oder mehrere Experimentierteilprozesse mit der damit verbundenen Vor- und Nachbereitung beinhalten konnte, auf der Ebene der Sichtstruktur gesucht und kodiert. Eine solche Einheit konnte entweder die gesamte Doppellektion betreffen oder aber nur einen Ausschnitt davon. Diese Einheit wurde in der Planung zusammen mit den kopierten Arbeitsmaterialien auf der Grundlage des Manuals ein erstes Mal qualitativ geratet. Im Anschluss daran wurden die beiden Videoaufnahmen der Front- und der Deckenkamera sowie die transkribierten Gesprächsinhalte für die zweite Ratingrunde beigezogen. Dies, um allenfalls die der Experimentiereinheit zugrundeliegende Idee besser erfassen zu können, Neues zu erfahren, was nicht in den Planungen schriftlich festgehalten worden war oder aber erklärende Gesten zum besseren Verständnis miteinbeziehen zu können. Dabei lag der Fokus der Ratinganalyse jeweils auf der zu Beginn des Gesprächs erfolgten Präsentation der Planung (nähere Informationen siehe Kap. I, 3.2.3). Dieses nach dem sogenannte Hypercodingverfahren nach Irion (2010) erfolgte Analyseverfahren zeichnet sich durch die synchrone Analyse und Einschätzung mehrerer Datenquellen aus. Dieses Verfahren eröffnet die Möglichkeit, Videodaten auch für die Rekonstruktion subjektiver Perspektiven zu nutzen (Irion, 2010, S. 142). Die Nutzung verschiedener Datenquellen ermöglicht gemäss Nijveldt (2007) und Stronge (2006) generell eine akurateren Erfassung der Performanz fachdidaktischen Wissens angehender Lehrpersonen. Auch Park & Suh (2015) treten für dieses Verfahren ein und begründen es mit der Komplexität des zu erfassenden Konstrukts. Die mit dieser Vorgehensweise erfassten Ratings wurden in einer vorgefertigten Excel-Datei festgehalten und in eine Gesamtdatei übertragen. Aufgrund der grossen Datenmenge und des damit verbundenen Aufwandes bei der Datenauswertung haben sich die Rater die Stichprobe ($n = 119$) etwa hälftig aufgeteilt.

3.2.5 Gütekriterien des Ratingmanuals

Die Qualität des Ratingmanuals lässt sich hauptsächlich mittels der drei Hauptgütekriterien, nämlich der Objektivität, Reliabilität und Validität bestimmen (Brückmann & Duit, 2014; Döring & Bortz, 2016). Dabei ist die Objektivität bei einem standardisierten Test vorauszusetzen, während die Reliabilität und Validität in der Regel anhand von empirisch bestimmten Kennwerten bzw. Koeffizienten geschätzt werden (Döring & Bortz, 2016, S. 464). Die Methodik zur Bestimmung der Objektivität, Reliabilität und Validität des Ratingmanuals wird in den nächstfolgenden Abschnitten beschrieben. Auch werden die in SPSS (IBM Corp., 2016) errechneten Ergebnisse zur Objektivität und Reliabilität berichtet. Ergebnisse der strukturellen Validierung des Ratingsmanuals werden auf der Basis der Forschungsfragen im Ergebnisteil der Teilstudie I präsentiert.

I. Objektivität

Die Objektivität eines Testinstruments bezieht sich auf die Beobachterunabhängigkeit der Ergebnisse. Die Anwendungs- und Auswertungsvorschriften dürfen nicht von der Person abhängen, die das Instrument anwendet, sondern müssen für die Skala eindeutig bzw. objektiv festgelegt sein (Döring & Bortz, 2016, S. 268).

Für die Durchführungsobjektivität wurde im Projekt KUBeX ein Manual entwickelt, in dem die Kriterien für die Beurteilung der Qualität fachdidaktischen Wissens in Unterrichtsplanungen für jeden Indikator der 5 Kategorien beschrieben und festgelegt worden sind. Das Ratinginstrument bildete die Grundlage für alle Ratings.

Die Einschätzung der Auswertungsobjektivität und damit der Zuverlässigkeit, der von zwei Ratern unabhängig voneinander gerateten Beobachtungen, ist mit Hilfe der Beurteilerübereinstimmung oder der Interrater-Reliabilität bestimmt worden (Hammann, Jördens, Krüger, Parchmann & Schecker, 2014). Mit der Beurteilerübereinstimmung wird der prozentuale Anteil der exakten Übereinstimmungen an der Gesamtzahl der Kodierungen geschätzt und damit die „Güte“ der Beobachtung stichprobentypisch berechnet. Die Übereinstimmung ist dabei „keine Eigenschaft der Ratingskala selbst [...], sondern [eine] Eigenschaft eines bestimmten Gebrauchs von Ratingskalen“ (Wirtz & Caspar, 2002, S. 24). Zum Nachweis der Interrater-Reliabilität und damit der Objektivität werden in der Regel 20% der videografierten Planungsgespräche doppelt kodiert (Brückmann & Duit, 2014). Die dafür verwendeten Videos werden zufällig aus dem Datensatz gezogen. Bei einem Vorgehen dieser Art wird angenommen, dass, sofern die Interrater-Reliabilität zufriedenstellende Werte erreicht, alle restlichen Videos in der gleichen Weise analysiert und geratet werden können (Brückmann und Duit, 2014). Die erste Einschätzung der Beurteilerübereinstimmung wird mit der prozentualen, exakten Übereinstimmung berechnet. Da bei diesem Berechnungsverfahren allerdings keinerlei gleiche, zufällig erfolgte Ratings mitberücksichtigt sind, werden deshalb für eine genauere Objektivitätsanalyse der Ratings weitere Analyseverfahren, wie das gewichtete Kappa für ordinale Daten und der Intraklassenkoeffizient (ICC) für metrische Daten empfohlen (Döring & Bortz, 2016). Sogenannte Ratingskalen, wie sie in der vorliegenden Studie eingesetzt worden sind (4-stufig von *trifft nicht zu* (1 Punkt) bis *trifft völlig zu* (4 Punkte)), werden in der Literatur oft als intervallskaliert angenommen (e.g. Döring & Bortz, 2016; Wirtz & Caspar, 2002, S. 33). Dies bedeutet, dass das Übereinstimmungsmass mittels Reliabilitätsmassen, wie dem gebräuchlichen ICC überprüft werden kann, so erfolgt in Arbeiten von Schulz (2011), Schmelzing (2010), Vogelsang (2014), Börlin (2012) und Goldston, Dantzer, Day und Webb (2013). Der ICC ist ein varianzanalytisches Mass für die Stärke des Zusammenhangs der Urteile zweier Beobachter,

die dieselben Ereignisse beobachtet und im Beobachtungsbogen beurteilt haben. Er wird auf der Datengrundlage bestehend aus kodierten Einzelrohdaten berechnet. Die *ICC*-Werte bewegen sich im Wertebereich von -1 bis +1. Je mehr sich diese dem Wert +1 nähern, umso höher ist auch die Reliabilität der Urteile (Döring & Bortz, 2016). *ICC*-Werte über .70 werden konventionell meist als gute Reliabilitätswerte gewertet (Döring & Bortz, 2016, S. 347; Wirtz & Caspar, 2002, S. 234). Entsprechend ist in diesem Fall ein Gruppenvergleich mittels der Rohdaten aussagekräftig (Wirtz & Caspar, 2002, S. 234). Allerdings kann der wahre Wert von Personen durch das Ratingurteil erst bei Reliabilitäten über .85 recht präzise eingegrenzt werden (Wirtz & Caspar, 2002, S. 234). Da bei sogenannten Rating-Skalen aber die Abstände zwischen den Skalenpunkten nicht mit Sicherheit gleich sind, empfehlen Wirtz und Caspar (2002) zusätzlich noch ein geeignetes Mass für ordinale Daten zu berechnen und anzugeben. Eine Möglichkeit stellt hier das gewichtete Kappa dar, das die Nähe der Übereinstimmungen berechnet und zufallskorrigiert ist (Döring & Bortz, 2016; Warrens, 2012). Kappa-Werte, die über .75 liegen, gelten nach konventionellen Standards als gut bis sehr gut, Werte zwischen .60 und .75 werden als zufriedenstellend eingestuft und Werte zwischen .40 und .60 als akzeptabel eingeordnet (Tab. 6).

Tabelle 6: Wertebereich und Interpretation von Cohens Kappa (Wirtz & Caspar, 2002)

Wertebereich von Cohens Kappa	Interpretation
.75 - 1.00	gute bis sehr gute Übereinstimmung
.60 - .75	zufriedenstellende Übereinstimmung
.40 - .60	akzeptable Übereinstimmung
-1.00 - .40	schlechte Übereinstimmung

Bei der Beurteilung der Kappa-Werte ist zu berücksichtigen, dass sie durch die Anzahl der Kategorien sowie der Kategorienbesetzungen (Indikatoren) beeinflusst werden. Eine inhaltlich differenzierte Interpretation wird daher empfohlen (Sim & Wright, 2005). Alternativ könnten auch Spearmans Rho oder Kendalls Tau, zwei Masse, welche die Stärke des ordinalen Zusammenhangs ausdrücken, ermittelt werden (Döring & Bortz, 2016, S. 347). Erweisen sich aber die Reliabilitäten beider Masse, d.h. für intervall- und auch ordinalskalierte Daten, als zufriedenstellend, so ist die Frage der Skalenniveaus für die Annahme der Reliabilität der Daten nicht kritisch (Wirtz & Caspar, 2002, S. 127).

Für die vorliegende Teilstudie I wurden insgesamt 23 zufällig ausgewählte Videos von zwei Schätzern unabhängig voneinander mit Hilfe des Ratingmanuals geratet. Die prozentuale, exakte Übereinstimmung beträgt 89%. Dies entspricht einer sehr hohen, aber noch ungenauen Einschätzung, da hier keine gleichen, zufällig erfolgten Ratings mitberücksichtigt worden sind. Aufgrund eines Entscheidungsschlüssels für die Wahl der *ICC*-Werte wurde weiter eine 2-faktorielle Varianzanalyse vorgenommen, da alle Fälle von beiden Ratern vollständig eingeschätzt worden sind (Schulz, 2011, S. 67). Die Berechnung der justierten *ICC*-Werte ist aus zwei Gründen erfolgt.

Tabelle 7: Statistische Kennwerte der Indikatoren des Ratingmanuals in der Übersicht (*ICC_{just}* zum experimentellen Handeln, 2 Rater 23 Videos)

Kategorien	Indikatoren	<i>ICC_{just}</i>	<i>F</i>	Konfidenzint.
Schülvorstellungen	Kenntnis	.79**	4.82	.51 - .91
	Erhebung	.83**	5.71	.61 - .92
	Umgang	.82**	5.45	.22 - .84
Fachliche Klärung	Fachliche Richtigkeit	.88**	8.83	.73 - .95
	Klärung relevanter Fachwörter	.96**	38.30	.94 - .99
	Material- und Geräteauswahl und -einsatz	.91**	11.18	.79 - .96
Unterrichtsziele	Zielauswahl	1.00		
	Zielorientierung	.98**	54.43	.95 - .99
	Zielklarheit	.99**	71.55	.98 - .99
Didaktische Strukturierung	Fachimmanente Einbettung	.95**	19.46	.88 - .98
	Lebensweltliche Einbettung	1.00		
	Kognitiver Anspruch	.96**	28.10	.91 - .98
	Sequenzierung der Lerninhalte	.92**	11.45	.80 - .96
	Zeitplanung	.91**	11.24	.80 - .96
Reflexivität	Ergebnisreflexion	.95**	22.79	.89 - .98
	Prozessreflexion	1.00		
	Begriffsbildung	.99**	72.46	.97 - .99
Exp. Handeln		.98**	88.33	.97 - .99

* $p < .05$, ** $p < .001$

Zum einen, weil sie auf der Abschätzung des Anteils von Varianz in den Messwerten bezogen auf die Varianz in und zwischen den Beurteilenden beruhen und damit die Strenge des Beurteilers miteinfliesst (Vogelsang, 2014, S. 330). Zum anderen, weil in dieser Studie die Beurteiler nicht zufällig ausgewählt worden sind, sondern bereits bei der Entwicklung des Manuals mitgewirkt haben (Schulz, 2011, S. 67). Die justierten ICC -Werte wurden pro Kategorie und pro Indikator berechnet (Tab. 7). Ausserdem wurde zur Absicherung das gewichtete Kappa ermittelt. Die ICC_{just} -Werte pro Kategorie sowie die gewichteten Kappa-Werte pro Indikator sind ergänzend im Anhang der Arbeit aufgeführt (cf. Anhang A-2, Tab. A2.2 & A2.3).

Die Analyse der ICC_{just} -Werte und damit der Übereinstimmungen beider Schätzer über das gesamte hoch inferente Ratingmanual zeigt insgesamt einen signifikanten und hohen Wert von .98. Die ICC_{just} -Werte pro Indikator liegen zwischen .79 und 1 und sind alle signifikant. Die unabhängige Bewertung beider Kodierer ist dabei bei 13 von 17 Indikatoren sehr hoch, bei 4 Indikatoren gut. Alle ICC_{just} -Werte sind mit Ausnahmen der Indikatoren der Kategorie Schülervorstellungen grösser als .85. Entsprechend sind auch die Konfidenzintervalle bei allen Indikatoren ausser den Indikatoren zu den Schülervorstellungen vergleichsweise eng. Das deutet auf präzisere Schätzungen hin, da das Konfidenzintervall einen Bereich beschreibt, in dem der errechnete ICC_{just} -Wert mit 95% Wahrscheinlichkeit liegt. Tabelle 7 zeigt ausserdem, dass bei den Indikatoren «Zielauswahl» (Kategorie Unterrichtsziele), «Lebensweltliche Einbettung» (Kategorie didaktische Strukturierung) und «Prozessreflexion» (Kategorie Reflexivität) ICC_{just} -Werte von 1 vorliegen. Dies bedeutet, dass hier die Ratings beider Rater vollständig übereingestimmt haben. Gemäss Wirtz und Caspar sind hier die Varianzen für beide Rater gleich, die Mittelwerte der Rater können allerdings verschieden sein (Wirtz & Caspar, 2002, S. 158). Dies lässt sich möglicherweise darauf zurückführen, dass die ersten beiden Indikatoren aus fachdidaktischer Sicht leichter zu beurteilen waren und der letzte Indikator selten oder nie aufgetreten ist.

Die Beurteilung der latenten Variablen «Fachdidaktisches Wissen zur Planung von Unterricht zum experimentellen Handeln» durch zwei Rater zeigt zusammengefasst eine sehr gute Übereinstimmung. Die beschriebenen Ergebnisse deuten darauf hin, dass die wahren Werte von Personen durch das Ratingurteil recht präzise eingegrenzt werden können. Diese Einschätzung deckt sich in etwa mit den Werten des gewichteten Cohens Kappa (cf. Anhang A-2, Tab. A 2.3). Die gewichteten Kappa-Werte für die Indikatoren sind stellenweise etwas tiefer und die Konfidenzintervalle bei gewissen Indikatoren grösser, aber gemäss Richtwerten immer noch im akzeptablen bis sehr guten Bereich.

Dies kann als Hinweis gedeutet werden, dass das Beobachtungssystem insgesamt problemlos auf die zu beobachtenden Fälle anwendbar ist und zu messgenauen Daten führt (cf. Döring & Bortz, 2016, S. 346).

Die Qualität eines psychometrischen Tests oder Fragebogens ist nicht zuletzt auch abhängig von der Art und der Zusammensetzung der Items (Döring & Bortz, 2016, S. 475). Dies gilt auch für Ratingmanuals, die den Schluss auf Tiefenstrukturen des Unterrichtens, die nicht direkt sichtbar sind, ermöglichen (Riegel & Macha, 2013). Entsprechend kann dies auch angenommen werden für die Analyse der vorliegenden Indikatoren, die das fachdidaktische Wissen zur Tiefenstruktur in Unterrichtsplanungen abbilden sollen. Eine Analyse der Items und die Berechnung der Reliabilität, d.h. der Messgenauigkeit des Instruments, sind daher unabdingbar (Bühner, 2011; Döring & Bortz, 2016; Moosbrugger & Kelava, 2012). Der nächste Abschnitt befasst sich mit der Itemanalyse und der Reliabilität des Ratinginstruments.

II. Itemanalyse und Reliabilität

Die Itemanalyse ist ein zentrales Instrument der klassischen Testtheorie, um die Qualität eines psychometrischen Tests, Fragebogens oder Ratinginstruments zu prüfen. Mit Hilfe der im Rahmen der Itemanalyse gewonnenen Ergebnisse ist eine genauere Qualitätsbeurteilung des gesamten „neuen“ Messinstruments möglich (Hartig et al., 2012). Daher werden im nächstfolgenden Abschnitt die Daten der 17 Items bzw. Indikatoren hinsichtlich ihrer relativen Häufigkeit (Rohwertverteilung) und ihrer zentralen Tendenz (Mittelwerte, Standardabweichung, Itemschwierigkeit, Itemtrennschärfe) analysiert. Auch wird die Reliabilität für das latente Konstrukt «Fachdidaktisches Wissen zur Planung von Unterricht zum experimentellen Handeln» ermittelt. Die Rohverteilung der Itemkennwerte ist Gegenstand des folgenden Abschnitts.

Rohwertverteilung der Indikatoren

Rohwertverteilungen der Indikatoren im Sinne ihrer relativen Häufigkeiten vermitteln einen ersten Überblick über die Verteilung der erfolgten Ratings, mit welchen die Indikatoren in den Planungen und Planungsgesprächen beobachtet werden konnten. Mit Hilfe von relativen Häufigkeitsverteilungen in Prozenten kann ermittelt werden, wie stark die Ergebnisse streuen und wie der gesamte Wertebereich ausgenutzt wird. Daraus lassen sich Schlüsse zur Art und Zusammensetzung der Indikatoren ziehen.

Eine erste grobe Analyse der vergebenen Codes der Indikatoren des Ratingmanuals in der Gesamtstichprobe ($N = 119$) zeigt, dass 49 Unterrichtsplanungen bei allen Indikatoren den Code 1 aufweisen. Das entspricht einer relativen Häufigkeit von 41% der Gesamt-

stichprobe (Tab. 8). Bei 70 Unterrichtsplanungen oder 59% aller erfassten Unterrichtsplanungen sind hingegen alle Codes (Code 1 - 4) vergeben worden, wenn auch in unterschiedlicher Häufigkeit. Das fachdidaktische Wissen zur Planung von Unterricht zum experimentellen Handeln konnte demzufolge nur in 70 der 119 Unterrichtsplanungen beobachtet werden, zumindest ansatzweise. Entsprechend kann die genauere Analyse der Ratings zu den Indikatoren nur auf der Basis der Teilstichprobe ($n = 70$) durchgeführt werden.

Tabelle 8: Absolute und relative Häufigkeiten (%) der vergebenen Codes

Ratings der Unterrichtsplanungen	absolute Häufigkeit $N = 119$	relative Häufigkeit (%) $N = 119$
nur Code 1	49	41
Code 1, 2, 3 & 4	70	59

In Abb. 20 ist die relative Häufigkeit der vergebenen Codes pro Indikator für die Teilstichprobe ($n = 70$) grafisch dargestellt.

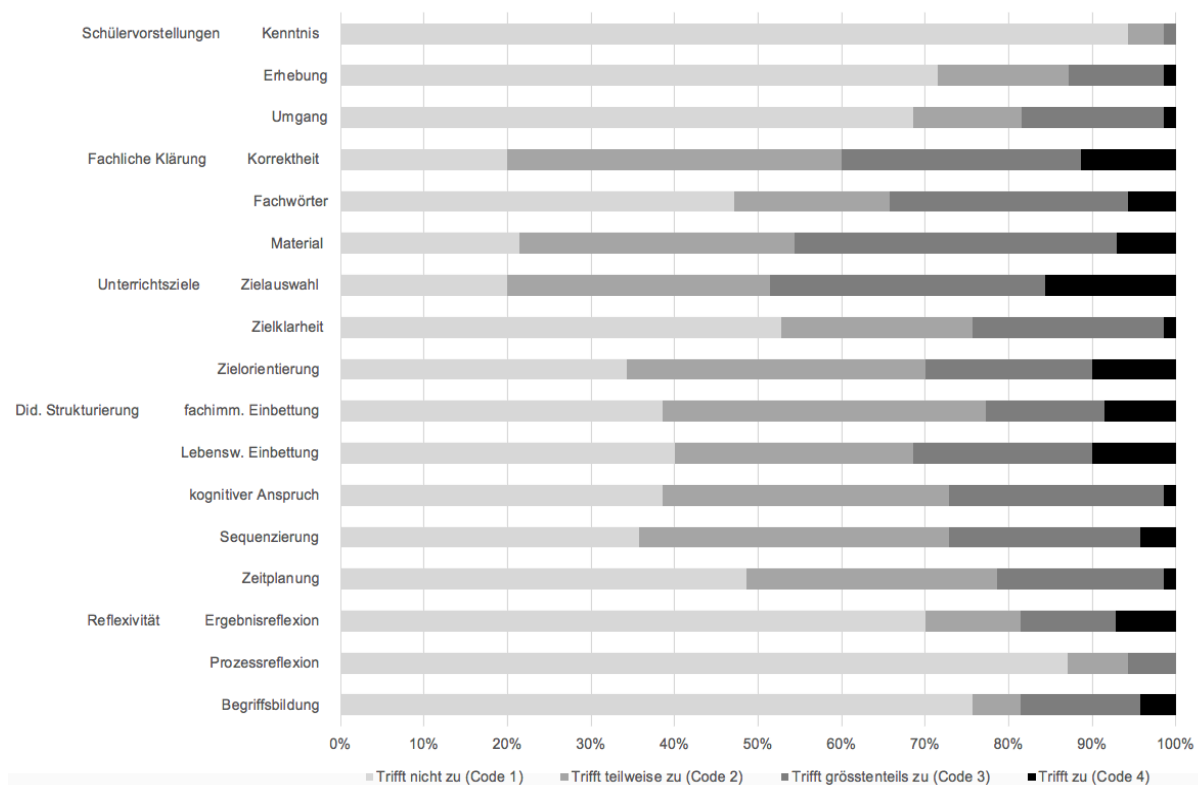


Abbildung 20: Relative Häufigkeiten der 4 Qualitätsstufen der 17 Indikatoren zum fachdidaktischen Wissen zur Planung von Unterricht zum experimentellen Handeln (Teilstichprobe $n = 70$)

Die Abbildung zeigt, dass die relativen Häufigkeiten des Codes 1 je nach Indikator teilweise sehr hoch sind. Sie liegen im Bereich zwischen 20% und 94.3%, d.h. sie streuen erheblich. Dabei fallen besonders die Kategorien «Schülervorstellungen» und «Reflexivität» mit ihren hohen Anteilen Code 1 auf. Diese liegen für die Kategorie «Schülervorstellungen» zwischen 68.6% und 94.3%, für die Kategorie «Reflexivität» zwischen 70.0% und 87.1%. Innerhalb dieser beiden Kategorien sind ausserdem keine höchsten Codes (Code = 4) bei den Indikatoren «Kenntnisse von Schülervorstellungen» (Kategorie «Schülervorstellungen») und «Prozessreflexion» (Kategorie «Reflexivität») vergeben worden. Ein etwas anderes Bild zeigen die relativen Häufigkeiten von Code 1 bei den Indikatoren der Kategorien «Fachliche Klärung», «Unterrichtsziele» und «Didaktische Strukturierung». Sie liegen zwischen 20.0% und 52.8%. Das bedeutet, dass die Ratings höherer Codes innerhalb dieser drei Kategorien insgesamt häufiger sind und die Rohwertverteilungen entsprechend besser ausfallen. Speziell anzumerken ist, dass die relative Häufigkeit des Codes 4 beim Indikator «Korrektheit» (Kategorie «Fachliche Klärung») 11.4% und beim Indikator «Zielauswahl» (Kategorie «Unterrichtsziele») 15.7% beträgt. Auch fällt auf, dass bei den Indikatoren «Relevante Fachwörter» sowie «Material- und Geräteauswahl» (Kategorie «Fachliche Klärung») Code 3 häufiger als Code 2 vergeben worden ist. Die prozentualen Unterschiede belaufen sich hier auf max. 10%.

Fasst man die wesentlichen Erkenntnisse aus der Rohwertverteilung zusammen, so zeigt sich, dass nur 70 der 119 Unterrichtsplanungen, -materialien und videografierten Planungsgespräche mit Codes 1 - 4 geratet worden sind. Die restlichen 49 weisen alle nur Code 1 auf. Auch die Rohwertverteilung der Teilstichprobe ($n = 70$) zeigt einen hohen Prozentsatz Code 1, insbesondere bei den Kategorien «Schülervorstellungen» und «Reflexivität». Die Indikatoren dieser beiden Kategorien sind entsprechend weniger häufig mit höheren Codes geratet worden. Die Kategorien «Fachliche Klärung», «Unterrichtsziele» und «Didaktische Strukturierung» hingegen sind im Vergleich dazu häufiger mit Code 2, 3, und 4 geratet worden, wobei je nach Indikator in unterschiedlicher Ausprägung. Entsprechend differenzieren diese Indikatoren stärker.

Ausgehend von dieser Analyse der Rohwertverteilungen werden in einem nächsten Schritt die Itemkennwerte der 17 Indikatoren für die Teilstichprobe ($n = 70$) ermittelt. Bei den Itemkennwerten handelt sich um deskriptive Grössen wie Mittelwert, Standardabweichung, Minimal- und Maximalwerte, Itemschwierigkeiten und -trennschärfen. Die Methodik der Berechnungen und die jeweiligen Ergebnisse werden im nächstfolgenden Abschnitt in dieser Reihenfolge vorgestellt.

Itemkennwerte der Indikatoren

Die Indikatoren eines Ratingmanuals lassen sich aufgrund ihrer unterschiedlichen Lösungs-raten, die in Form von Itemschwierigkeiten (Itemschwierigkeitsindizes) quantifizierbar sind, näher beschreiben (Döring & Bortz, 2016). Im Fall von Ratingskalen, die zwar oft als intervallskaliert angenommen werden, meist aber Ordinalniveau besitzen, wäre es im Grunde angemessen, den Median als Schwierigkeitsindex zu ermitteln (Bühner, 2011). Da die Ratingskalen aber häufig wenige Antwortkategorien aufweisen, und damit der Median für viele Indikatoren ähnlich ausfällt, ist die Differenzierungsfähigkeit damit eingeschränkt. Bühner (2011) empfiehlt daher die Berechnung der Mittelwerte als Schwierigkeitsindex. Die Mittelwerte sind für das vorliegende Ratingmanual ermittelt worden. Diese Berechnungen wurden zusammen mit den Berechnungen weiterer Itemkennwerte in SPSS 24 durchgeführt (IBM Corp., 2016). Der durchschnittliche *Mittelwert* über alle 17 Indikatoren beträgt für die Teilstichprobe ($n = 70$) 1.80 ($SD = 0.39$). Die Mittelwerte der einzelnen Indikatoren liegen zwischen 1.07 und 2.44 (Tab. 9).

Tabelle 9: Itemkennwerte der 17 Indikatoren des Ratingmanuals ($n = 70$)

Kategorien	Indikator	<i>M</i>	<i>SD</i>	<i>min.</i>	<i>max.</i>	<i>r_{it}</i>	<i>P</i> (%)
Schüler- vorstellungen	Kenntnis	1.07	0.31	1	3	.12	2.38
	Erhebung	1.43	0.75	1	4	.60	33.33
	Umgang	1.51	0.83	1	4	.71	17.14
Fachliche Klärung	Fachliche Richtigkeit	2.31	0.93	1	4	.72	43.81
	Klärung relevanter Fachwörter	1.93	0.99	1	4	.75	30.95
	Material- und Geräteauswahl und -einsatz	2.31	0.89	1	4	.63	43.81
Unterrichtsziele	Zielauswahl	2.44	0.99	1	4	.79	30.95
	Zielorientierung	1.73	0.87	1	4	.74	33.81
	Zielklarheit	2.06	0.98	1	4	.89	30.00
Didaktische Strukturierung	Fachimmanente Einbettung	1.93	0.94	1	4	.82	31.90
	Lebensweltliche Einbettung	2.01	1.01	1	4	.31	24.76
	Kognitiver Anspruch	1.90	0.84	1	4	.74	48.10
	Sequenzierung der Lerninhalte	1.96	0.88	1	4	.85	24.29
	Zeitplanung	1.74	0.83	1	4	.36	24.29
	Ergebnisreflexion	1.56	0.96	1	4	.64	18.57
Reflexivität	Prozessreflexion	1.19	0.52	1	3	.43	6.19
	Begriffsbildung	1.47	0.90	1	4	.77	15.71

Bei der vorliegenden 4-stufigen Ratingskala von 1 (*trifft nicht zu*) bis 4 Punkten (*trifft völlig zu*) und einer Skalenmitte von 2.5 bedeutet dies, dass die Indikatoren des Ratingmanuals testtheoretisch als anspruchsvoll einzustufen und im Schnitt nicht mit sehr hoher Qualität erfasst worden sind. Die Indikatoren der Kategorien «Schülervorstellungen» und «Reflexivität» weisen dabei die tiefsten Mittelwerte auf, während die Mittelwerte der Indikatoren der Kategorien «Fachliche Klärung», «Unterrichtsziele» sowie «Didaktische Strukturierung» vergleichsweise höher liegen.

Die *Standardabweichungen*, die ein Mass für die Streubreite der Werte eines Merkmals rund um deren Mittelwert darstellen, zeigen hierzu noch ein etwas differenzierteres Bild. Die Standardabweichungen aller Indikatoren bewegen sich zwischen 0.31 und 1.01. Während die Indikatoren «Kenntnis» (Kategorie «Schülervorstellungen») und «Prozessreflexion» (Kategorie «Reflexivität») mit 0.31 bzw. 0.52 die tiefsten Standardabweichungen und damit auch geringe Varianzen aufweisen, sind die Streuung und damit die Varianz bei den übrigen Indikatoren vergleichsweise grösser. Insbesondere die Indikatoren der Kategorien «Fachliche Klärung», «Unterrichtsziele» sowie «Didaktische Strukturierung» weisen höhere Standardabweichungen zwischen .83 und 1.01 auf. Inhaltlich bedeutet dies, dass sich die Qualitäten eines Indikators bei verschiedenen Personen klarer unterscheiden (cf. Bühner, 2011, S. 259).

Die *Trennschärfen*, die ein Mass dafür sind, wie gut ein einzelnes Item das Zielkonstrukt des Tests misst, können Werte zwischen -1 und +1 annehmen (Hartig et al., 2012). Die Trennschärfen der 17 Indikatoren liegen mit einer Ausnahme zwischen .31 und .89. Sie sind damit $> .30$ und gemäss Bühner (2011, S. 81) im mittleren (.30 - .50) bis hohen Bereich ($> .50$) anzusiedeln. Dies bedeutet, dass die einzelnen Indikatoren das Ziel- bzw. Kompetenzkonstrukt «Fachdidaktisches Wissen zur Planung von Unterricht zum experimentellen Handeln» gut messen. Je höher die Werte der einzelnen Indikatoren nämlich ausfallen, desto ähnlicher differenzieren sie wie der Gesamttest (Hartig et al., 2012). Allerdings liegt die Trennschärfe des Indikators «Kenntnis» aus der Kategorie «Schülervorstellungen» mit .12 unter 30. Es muss daher geprüft werden, ob dieser Indikator tatsächlich für die Ermittlung der strukturellen Validität (cf. Kap. I, 3.2.5) berücksichtigt werden kann. Bühner (2011) erwähnt hierzu aber, dass „niedrige Trennschärfen nicht zwangsweise zur Entfernung der Items aus dem Test führen sollten“ (Bühner, 2011, S. 81).

Zur weiteren Beschreibung der Indikatoren lassen sich auch Zusammenhänge zwischen gewissen Kennwerten, wie u.a. der Standardabweichung und der Trennschärfe pro Indikator, prüfen (Bühner, 2011, S. 259). Bei den vorliegenden Daten korreliert die Höhe der Standardabweichungen mit den Trennschärfen der Indikatoren. Die Korrelation beträgt .66 und ist auf dem Niveau von .01 zweiseitig signifikant. Daraus kann abgeleitet werden, dass

es sich bei den zu verzeichnenden Unterschieden der Indikatoren um systematische und nicht um durch Messfehler bedingte Unterschiede handelt (Bühner, 2011, S. 259).

Die Analyse der Itemkennwerte hat ergeben, dass die Indikatoren des Ratingmanuals anspruchsvoll sind. Vor allem die Kategorien «Schülervorstellungen» und «Reflexivität» weisen vergleichsweise tiefere Mittelwerte mit z.T. geringeren Standardabweichungen auf, während die anderen Kategorien besser abschneiden. Die Trennschärfen der einzelnen Indikatoren sind bis auf den Indikator «Kenntnis» aus der Kategorie «Schülervorstellungen» grösser .30 und korrelieren signifikant mit den Standardabweichungen. Die Indikatoren bilden damit das latente Konstrukt «Fachdidaktisches Wissen zur Planung von Unterricht zum experimentellen Handeln» gut ab und die Unterschiede sind nicht auf Messfehler zurückzuführen. Aus Sicht der Objektivität lässt sich das Ratingmanual demzufolge insgesamt als Skala einsetzen. Der Indikator «Kenntnis» aus der Kategorie «Schülervorstellungen» muss allerdings in Bezug auf seine Eignung weiter geprüft werden.

Die Frage, in welcher Genauigkeit das Ratingmanual das Konstrukt «Fachdidaktisches Wissen zur Planung von Unterricht zum experimentellen Handeln» misst, ist damit jedoch nicht geklärt. Dies lässt sich über die Reliabilität prüfen. Der folgende Abschnitt geht auf diesen Aspekt der Itemanalyse näher ein.

Reliabilität des Ratingmanuals

Besteht ein Ratingverfahren aus Indikatoren, die alle das gleiche Merkmal erfassen und damit eine Skala bilden, so kann aufgrund der Zusammenhangsstrukturen der Indikatoren auf die interne Konsistenz geschlossen und die Reliabilität des Verfahrens geschätzt werden. Mit diesem Verfahren wird geprüft, inwieweit ähnliche Aufgaben hinsichtlich der Qualität ihrer Bearbeitung ähnlich gelöst werden (Schecker, 2014). Die Reliabilität des Verfahrens kennzeichnet dabei den Grad der Genauigkeit bzw. Zuverlässigkeit mit dem das geprüfte Merkmal gemessen wird und – im Sinne der klassischen Testtheorie – den Grad der Messfehlerfreiheit. Allgemein lässt sich die Reliabilität als Anteil der wahren Varianz an der beobachteten Varianz definieren (Döring & Bortz, 2016, S. 465).

Das mit Abstand gebräuchlichste quantitative Mass für die Bestimmung der Reliabilität stellt die interne Konsistenz dar. Diese wird typischerweise mit Hilfe von Cronbachs Alpha-Koeffizienten berechnet. Dieser Koeffizient lässt sich über die Varianzen der Items oder über die mittlere Interkorrelation zwischen den Items berechnen (Schecker, 2014). Dabei ist die interne Konsistenz eines Tests umso höher, je höher die Korrelationen zwischen den Items im Durchschnitt sind. Der Wertebereich von Cronbachs α bewegt sich zwischen 0 (Messung völlig unzuverlässig, nur aus zufälligen Messfehlern bestehend) und

1 (Messung perfekt zuverlässig, ohne jegliche Beeinträchtigung durch Messfehler). Durch Eliminierung von Items mit geringer Trennschärfe kann der Reliabilitätskoeffizient gegebenenfalls auch erhöht werden (Döring & Bortz, 2016, S. 271). Die Cronbachs α -Werte sind darüber hinaus auch von der Anzahl der Items abhängig. Ein Test mit vielen Items kann eine hohe konsistenzanalytische Reliabilitätsschätzung erzielen, auch wenn mit diesem Test tatsächlich zwei oder mehrere voneinander unabhängige Dimensionen erfasst werden (Schermelleh-Engel & Werner, 2012, S. 133). Ein hoher Cronbachs α -Wert ist somit ein Mass für die interne Konsistenz, aber kein Mass dafür, ob die Skala tatsächlich nur ein Merkmal erfasst. Dies bedeutet, dass bei der Interpretation der Höhe von Reliabilitätskoeffizienten immer auch die Beschaffenheit der Kategorien sowie die Anzahl der Merkmalsausprägungen berücksichtigt werden sollten. Die Mehrdimensionalität einer Skala kann über sogenannte Faktorenanalysen geprüft werden (Schecker, 2014).

Die Reliabilität des Ratingmanuals der vorliegenden Studie beläuft sich für die Teilstichprobe $n = 70$ auf .95. In der Literatur findet sich ein breiter Diskurs zu den Richtwerten von Cronbachs α bezüglich der Aussagekraft für die Messgenauigkeit. Gemäss Döring & Bortz (2016, S. 443), die sich u.a. auch auf Bühner (2011, S. 81) beziehen, werden Reliabilitätswerte $> .90$ üblicherweise als hoch und Werte $> .80$ als ausreichend eingestuft. Kline (2016, S. 92) erachtet demgegenüber Cronbachs α -Werte $> .90$ als exzellent, Werte $> .80$ als gut und Werte $> .70$ als adäquat. Moosbrugger & Kelava (2012) geben an, dass der Reliabilitätskoeffizient eines guten Tests .70 nicht unterschreiten sollte. Field (2013) merkt an, dass eine Reliabilität $> .60$ bei einer geringen Anzahl Items eine adäquate interne Konsistenz darstellt. Ausgehend von diesen unterschiedlich interpretierten Richtwerten, kann festgehalten werden, dass die Reliabilität und damit die Messgenauigkeit des Gesamtkonstrukts «Fachdidaktisches Wissen zur Planung von Unterricht zum experimentellen Handeln» insgesamt sehr hoch ist und die Indikatoren das Gesamtkonstrukt repräsentieren. Das im Rahmen des Projekts KUBeX entwickelte Ratingmanual, bestehend aus 5 Kategorien und 17 Items bzw. Indikatoren, kann demzufolge auch aus Sicht der Reliabilität als Skala verwendet werden. Die Frage der Dimensionalität und in diesem Zusammenhang der Validität des Erhebungsinstruments ist aber noch offen. In den nächstfolgenden zwei Abschnitten werden vor dem Hintergrund der Forschungsfragen der Teilstudie I (Kap. I, 2.11) zum einen die Methodik zur Untersuchung der Validität, aber auch die statistischen Verfahren näher erläutert.

III. Validität

Wie vorangehend berichtet, kann aufgrund einer hohen Reliabilität nicht auf die Dimensionalität und damit auf einen der Aspekte der Validität eines Konstruktes geschlossen

werden. Die Einschätzung der Qualität eines Erhebungsinstruments, mit dem Rückschlüsse auf die Kompetenz vorgenommen werden, die nicht direkt beobachtet werden kann, ist daher von grosser Bedeutung (Roick & Henschel, 2015). Aus diesem Grund wird die Validität auch „als das grundlegendste der drei Hauptgütekriterien wissenschaftlicher Tests angesehen“ (Schmiemann & Lücken, 2014). Dies gilt insbesondere auch für *performance assessments* (Messick, 1995) wie Videostudien (Brückmann & Duit, 2014) und für die damit verbundene Analyse von Kompetenzmodellen, die zunächst nur hypothetische Beschreibungsmodelle darstellen, die auf ihre Gültigkeit zu überprüfen sind (Schaper, 2009; cf. Kap. I, 2.4.2)

Für eine angemessene Beurteilung der Testqualität gibt es keine standardisierten Verfahren, weil die Validität selbst ein komplexes theoretisches Gebilde darstellt. Dies widerspiegelt sich auch im regen Diskurs in der Literatur hinsichtlich der Aspekte, die die Validität definieren (Kane, 2013; Leutner, Hartig & Jude, 2008). Klassischerweise wird die Validität mit dem Grad der Genauigkeit umschrieben, mit der ein Erhebungsinstrument das latente Merkmal misst (Döring & Bortz, 2016). Nach diesem Verständnis beschränkt sich die Validität nur auf die statistisch quantifizierbare Eigenschaft eines Erhebungsinstruments. Solche Kennwerte sagen aber wenig über die Angemessenheit eines Instruments aus (Schaper, 2014). Gemäss Messick (1995) und Kane (2013) sollte sich die Validierung einer Messung daher nicht allein auf den numerischen Koeffizienten stützen, sondern es sollten auch die aktuellen und potenziellen Konsequenzen der Interpretation und Nutzung des Instruments in die Analyse einbezogen werden. Validität wird in diesem Sinn als ein breit definiertes Gütekriterium verstanden, das sich auf verschiedene Qualitätsaspekte eines Tests bezieht (Hartig et al., 2012). Dieses erweiterte Verständnis der Validität und in diesem Zusammenhang auch der Validierung nimmt somit Bezug auf die dem Testverfahren zugrunde liegenden theoretischen Annahmen sowie auch auf mögliche Interpretationen und Schlussfolgerungen der Ergebnisse (Messick, 1995). Für die Einschätzung der Validität eines Verfahrens sind daher neben theoretisch-rationalen Begründungen und Prinzipien auch empirische Evidenzen bzw. Erkenntnisse zu ermitteln, mit welchen die Interpretation von Testergebnissen und die daraus gezogenen Schlussfolgerungen begründet und plausibel gestützt werden können (Messick, 1995, Schaper, 2014). Leuders fasst Messicks elaboriertes Validitätskonzept wie folgt zusammen: „Validität [...] [ist] die Gesamtbewertung der theoretischen Argumente und empirischen Evidenzen für die Angemessenheit einer Leistungsmessung (Assessment) und zwar sowohl ihrer Interpretation als auch der Konsequenzen ihrer Anwendung“ (Leuders, 2014, S. 11). Dieser sogenannte „*argument based approach to validation*“ lässt sich sinnvollerweise auch auf die Validierung von Kompetenzmodellen übertragen und wird als kontinuierlicher argumentativer Prozess aufgefasst, der niemals

abgeschlossen ist (Cauet, 2016, S. 37; Schaper, 2014, S. 23). Kane schreibt dazu: “Validity is a matter of degree, and it may change over time as the interpretations/uses develop and as new evidence accumulates” (Kane, 2013, S. 3).

Sechs Aspekte von Validität

Nach der klassischen Einteilung von Validität wird zwischen Inhalts-, Kriteriums- und Konstruktvalidität unterschieden (American Psychological Association, [APA], 1954). Diese drei Validitätsarten hat Messick in seinem Modell (Messick, 1989) zu einem umfassenderen Validitätsverständnis weiterentwickelt und insgesamt sechs Arten von Evidenzquellen in Form von Validitätsaspekten beschrieben, auf die auch die aktuellen Standards der APA (2002) für psychometrische Tests zurückgreifen (Wilhelm & Kunina-Habenicht, 2015). Es sind dies die inhaltliche, die kognitive und die strukturelle Validität sowie die Verallgemeinerbarkeit, die externe und die konsequentielle Validität (Tab. 10).

Tabelle 10: Validitätsaspekte nach Messick (1995)

Validitätsaspekt	Beschreibung
Inhaltliche Validität	Curriculare und theoretische Absicherung des modellierten Bereichs (content aspect)
Kognitive Validität	Passung der kognitiven Prozesse bei der Kompetenzerfassung zum postulierten theoretischen Kompetenzmodell (substantive aspect)
Strukturelle Validität	Passung von theoretischem Kompetenzmodell und gewählttem psychometrischen Messmodell (structural aspect)
Verallgemeinerbarkeit	Angemessenheit einer über die Aufgaben- und Personengruppe hinausgehenden Interpretation (generalizability aspect)
Externe Validität	Angemessenheit mit Blick auf konvergente, diskriminante und prädiktive Zusammenhänge mit anderen Konstrukten (external aspect)
Konsequentielle Validität	Angemessenheit der Nutzung im pädagogischen und bildungspolitischen Kontext (consequential aspect)

Diese sechs Validitätsaspekte, insbesondere die ersten drei, die angesichts der Ausgangslage der vorliegenden Arbeit im Zentrum stehen, werden in Anlehnung an Leuders (2014) und Schaper (2014) in der Folge beschrieben.

Die *inhaltliche Validität* prüft die bei der Kompetenzmodellierung erfolgende Setzung des inhaltlichen Rahmens, der in seiner Breite und seinem Auflösungsvermögen erheblich variieren kann. Es geht dabei um eine curriculare und theoretische Absicherung des inhaltlichen Rahmens. Dieser kann zum einen sehr breit angelegt sein und die gesamte berufliche

Domäne umfassen. Zum anderen kann er auch nur Teilbereiche oder inhaltliche Dimensionen der Domäne, d.h. abgrenzbare Aktivitäten bzw. Tätigkeiten, einschliessen (Schaper et al., 2009). Normativ geprägte, breite Rahmenkonzepte lassen sich in der Regel mit Hilfe konsensueller Verständigung unter Expertinnen und Experten entwickeln und abstützen (Schaper, 2014). Standardorientierte bzw. professionswissensorientierte Modelle professioneller Handlungskompetenz können hier u.a. zugeordnet werden (cf. Kap. I, 2.4.2). Enger gefasste Rahmenkonzepte, die z.B. das fachdidaktische Wissen betreffen, beziehen sich in der Regel auf Theorien über die jeweils relevanten Kognitionen. Die inhaltliche Modellierung kann dabei sowohl theoretisch mit Hilfe von Expertinnen und Experten als auch empirisch über die Analyse individueller Lernprozesse erfolgen (Leuders, 2014). Modelle zur Erfassung fachdidaktischen Wissens wären beispielsweise hier einzuordnen (cf. Kap. I, 2.5.2). Erfüllt ist die inhaltliche Validität jeweils dann, wenn im Rahmen der Operationalisierung eine gründliche Konzeptspezifikation und eine sorgfältige theoretische Ableitung, Definition und Begründung der Kategorien sichergestellt wird (Döring & Bortz, 2016, S. 347) und die Frage beantwortet werden kann, ob und in welcher Weise die im Modell enthaltenen Items bzw. Indikatoren die Kompetenz qualitativ abbilden (Schaper, 2009).

Die *kognitive Validität* konkretisiert die Theorieelemente anhand von klar umschriebenen Aufgabensituationen. Damit wird eine Verbindung zwischen den theoretischen Kompetenzmodellen und ihrer empirischen Erfassung hergestellt und die Passung der kognitiven Prozesse bei der Kompetenzerfassung zum postulierten theoretischen Kompetenzmodell geprüft. Je nach diagnostischem Format werden dabei unterschiedliche Bearbeitungsformen derselben Aufgabe angeregt. Entsprechend kann die Operationalisierung eher distal oder proximal erfolgen. Eine möglichst situationsnahe Erfassung stellt für Shavelson (2010, S. 46) eine von vier zentralen Aspekten der Qualität von diagnostischen Formaten bzw. von Aufgaben dar. Diese postulierten Qualitätskriterien lauten:

1. Diagnostische Aufgaben haben einen Bezug zur realen Welt und sind möglichst situationsnah.
2. Diagnostische Aufgaben erfordern die Anwendung komplexer Fähigkeiten und Fertigkeiten
3. Diagnostische Aufgaben sind für die Praxis zugänglich.
4. Diagnostische Aufgaben sind in standardisierten Situationen umsetzbar.

Die Prüfung kognitiver Validität beruht ebenfalls auf theoretisch wie auf empirisch fundierten Argumenten. Diese können mit Hilfe von Expertinnen und Experten generiert werden. Nachweise können aber auch durch eine Untersuchung der Bearbeitungsprozesse unter testnahen Bedingungen, beispielsweise mit Hilfe der Methode des lauten Denkens oder

Stimulated Recall-Techniken, erzielt werden. Solche Analysen tragen dazu bei, grundlegende Schwierigkeiten der Studierenden bei der Lösung der Aufgabe aufzudecken.

Unter *struktureller Validität* wird gemäss Messick (1995) die Passung von theoretischem Kompetenzmodell und gewählttem psychometrischen Messmodell verstanden. Es wird überprüft, „ob das bei einer Messung zugrundeliegende implizite oder explizite Messmodell (*scoring model*) mit den Strukturen des Konstrukts übereinstimmt“ (Leuders, 2014, S. 15). In der klassischen Testtheorie wird ermittelt, ob die Anzahl der Variablen und in diesem Zusammenhang die Dimensionalität des Messmodells und die Bewertung und Gewichtung der Lösungen kongruent, d.h. valide sind. Dabei wird zwischen konvergenter, diskriminanter und faktorieller Validität unterschieden (Döring & Bortz, 2016). Konvergente Validität bedeutet, dass der zu validierende Test mit eng verwandten Konstrukten hoch korreliert. Diskriminante Validität liegt vor, wenn der zu validierende Test mit weit entfernten Konstrukten oder konstruktfernen Verfahren nicht oder wenig korreliert. Bei der faktoriellen Validität werden die Dimensionen eines Konstruktes überprüft. Sie liegt vor, „wenn bei einem mehrdimensionalen Test sich die inhaltlich zu einer Subdimension gehörenden Items jeweils auch empirisch zu einem Faktor bündeln lassen“ (Döring & Bortz, 2016, S. 446). Im Rahmen der Item-Response-Theorie bezieht sich die Frage der strukturellen Validität im Vergleich dazu auf ein ausgewähltes probabilistisches Messmodell. Ein solches Messmodell beschreibt „das Verhalten (Response) von Individuen bei bestimmten Situationen (Aufgaben, Items) in probabilistischer Abhängigkeit von Anforderungsmerkmalen der Situationen (Itemmerkmale) und Dispositionen der Personen (latente Personenmerkmale)“ (Leuders, 2014, S. 17). Die Lokalisation sowohl von Aufgabenschwierigkeiten als auch Personenmerkmalen auf einer gemeinsamen Skala wird damit möglich (Hartig & Frey, 2013). Die so ermittelte strukturelle Validität ist somit eine theoriegeleitete Entscheidung. Diese kann argumentativ aufgrund des Zusammenspiels der Skalenqualität mit dem Wissen bezüglich kognitiver Prozesse bei der Aufgabenlösung erfolgen, oder aber auch komparativ auf der Basis eines Vergleichs der Passung verschiedener konkurrierender Modelle mit den empirischen Daten (Schaper, 2014, S. 33).

Der Validitätsaspekt der *Verallgemeinerbarkeit* bezieht sich auf die Angemessenheit einer über die Aufgaben- und Personengruppe hinausgehenden Interpretation. Dabei sollte die Konstruktion eines Messmodells, das für einen bestimmten Kompetenzbereich allgemeingültig ist und mit dem Aussagen gemacht werden sollen, nicht von bestimmten Bedingungen (Stichprobe, Auswahl der Items, Verfahren der Antwortbewertung) des Konstruktionsprozesses abhängen. Mittels der sogenannten Generalisierbarkeitstheorie (Shavelson & Webb, 2006) lassen sich in der klassischen Testtheorie unerwünschte Varianzen ermitteln

und damit empirische Argumente für die Verallgemeinerbarkeit des Kompetenzmodells oder Konstruktes gewinnen.

Bei der *externen Validität* werden hypothesengeleitete Zusammenhänge zwischen Testergebnissen und Aussenkriterien (manifesten Merkmalen), die für diagnostische Entscheidungen wichtig sind, geprüft. Die ermittelten Zusammenhänge zwischen der Messung und externen Kriterien sollten mit den erwarteten Zusammenhängen übereinstimmen (Kane, 2006). Je nach Domäne oder Konstrukt werden dabei unterschiedliche Aussenkriterien beigezogen, in der Lehrpersonenbildungsforschung beispielsweise häufig Schulnoten. Diese Form der Validierung entspricht damit der klassischen Kriteriumsvalidität. Mit Hilfe dieses Validitätsaspektes kann von einem Testergebnis auf ein für diagnostische bzw. prognostische Entscheidungen relevantes Kriterium ausserhalb der Testsituation geschlossen werden (Hartig et al., 2012; Schaper, 2014).

Die *konsequentielle Validität* überprüft die Angemessenheit der Nutzung von Kompetenzmodellen im pädagogischen und bildungspolitischen Kontext, da Kompetenzmodellierungen zunehmend auch für die Erhebung und Rückmeldung von Leistungen in Bereichen des Schulsystems eingesetzt werden. Schaper (2014) und Leuders (2014) erwähnen hierzu Bereiche wie (a) die Bildungsmonitorings auf Systemebene, (b) die Rückmeldungen zu Leistungen auf Klassen- oder Schulebene, (c) die Professionalisierung im Sinne diagnostischer Kompetenzen von Lehrpersonen oder aber auch (d) die Unterstützung individueller Diagnose zur Vorbereitung pädagogischer Förderentscheidungen.

Validierung des Ratinginstruments aus dem Projekt KUBeX

Aufgrund des explorativen Charakters der vorliegenden Studie sind für die Validierung des Ratinginstruments die ersten drei Validitätsaspekte nach Messick (1995) von Bedeutung, nämlich die inhaltliche, die kognitive und die strukturelle Validität. Die ersten beiden Validitätsaspekte werden in der Folge besprochen, der Aspekt der strukturellen Validität ist Gegenstand des Ergebniskapitels (Kap. I, 4).

Das im Projekt KUBeX entwickelte Ratingmanual ist auf einen engen inhaltlichen Rahmen ausgerichtet, nämlich «Fachdidaktisches Wissen zur Planung von Unterricht zum experimentellen Handeln». Die inhaltliche Modellierung dieses Instruments ist dabei einerseits auf der Basis von theoretischen wie auch empirischen Ergebnissen zu relevanten Kognitionen erfolgt, die für die Planung von qualitativem und lernwirksamem Unterricht erforderlich sind. Andererseits sind auch Erkenntnisse aus Austauschen mit Expertinnen und Experten im Modell berücksichtigt worden. Strukturbildend für das Ratingmanual waren vor allem die Kategorien des in der Naturwissenschaftsdidaktik anerkannten Planungs-

modells der «Didaktischen Rekonstruktion» sowie die damit verbundenen Wissensfacetten fachdidaktischen Wissens (Kap. I, 2.5.2). Ergänzend dazu ist auch die Reflexivität, ein wesentliches Kennzeichen einer holistischen Auffassung experimentellen Handelns, im Strukturmodell als Kategorie berücksichtigt worden. Aufgrund dieser Vorgehensweise und der damit verbundenen Konzeptspezifikationen, die in Kap. I, 3.2.4 genau beschrieben sind, kann das Kriterium der *inhaltlichen Validität* als erfüllt betrachtet werden.

Die *kognitive Validität* des Ratingmanuals, welche die Passung der kognitiven Prozesse bei der Kompetenzerfassung zum postulierten theoretischen Kompetenzmodell prüft, bedarf in der vorliegenden Studie einer differenzierten Betrachtung des Erhebungsdesigns. Mit einer konkreten Planungsaufgabe, wie sie im Schulalltag erfolgen kann, ist «Fachdidaktisches Wissen zur Planung von Unterricht zum experimentellen Handeln» im Projekt KUBeX proximal und damit auch ökologisch valider erfasst worden. Damit wurde die Handlungsebene berücksichtigt, die die Handlungsrelevanz des erfassten Professionswissens letztendlich erst begründbar macht (Vogelsang & Reinhold, 2013a). Der Planungsauftrag wies mit dem experimentellen Handeln als Fachinhalt und dem biologischen Inhalt «Visuelle Wahrnehmung» einen Bezug zur realen Welt auf, da ersteres in den Bildungsstandards verankert (EDK, 2011; KMK, 2005) und letzteres Teil der Curricula der Sekundarstufe I ist. Ausserdem erforderte das Lösen des Planungsauftrags komplexe Fähigkeiten und Fertigkeiten und war entsprechend praxisnah. Die Kompetenz wurde zudem in einer standardisierten Situation erhoben (cf. Kap. I, 3.2.3). So gesehen erfüllte der Planungsauftrag als diagnostische Aufgabe alle vier Qualitätskriterien nach Shavelson (2010). Auch wurden die nach Oser, Heinzer und Salzmann (2010) wesentlichen Aspekte des Lehrerberufs wie Situativität, Authentizität, Komplexität sowie Kontextgebundenheit des unterrichtlichen Handelns bei der Kompetenzdiagnostik grösstenteils berücksichtigt. Allerdings muss kritisch angemerkt werden, dass aufgrund der nicht vollständig randomisierten Zuteilung der Stichprobe in Interventions- und Kontrollgruppe und damit des quasiexperimentellen Designs die Validität geschmälert ist. Ausserdem beruhte der Planungsauftrag auf einer fiktiven Situation und die Studierenden mussten die Unterrichtsplanungen aufgrund einer Vignette (cf. Anhang A-1, Abb. A1.1) herstellen. Die Erhebung war zudem in die laufende Ausbildung angehender Sekundarlehrpersonen eingebettet, was zur Folge hatte, dass der Planungsauftrag zusätzlich zu anderen Aufträgen in den jeweiligen Modulen geleistet werden musste. Auch war eine Umsetzung der Planung nicht möglich. Um sicherzustellen, dass die Ergebnisse nicht durch diesen Kontext beeinflusst worden sind, wäre mit Hilfe der Methode des lauten Denkens oder *Stimulated Recall*-Techniken zu prüfen gewesen, wie die Planungsprozesse unter diesen Bedingungen tatsächlich erfolgt wären. Vor diesem Hintergrund kann die kognitive Validität des Instruments nicht vollständig als erfüllt eingestuft werden.

Die *strukturelle Validität* des kategoriengeleiteten Ratingmanuals zur qualitativen Einschätzung des fachdidaktischen Wissens zur Planung von Unterricht zum experimentellen Handeln wird im folgenden Kapitel entlang der Forschungsfragen der Teilstudie I dargestellt und erläutert (Kap. I, 4). Die Untersuchung erfolgt mittels Faktorenanalyse, mit der sich gleichzeitig auch klären lässt, ob die hohen Cronbachs-Alpha-Werte nicht auf inhaltliche Redundanzen oder gar auf ein zu enges Konstrukt zurückzuführen sind. Die statistischen Verfahren, die zur Beantwortung der Forschungsfragen beigezogen worden sind und mit deren Hilfe die Dimensionalität des Konstruktes geprüft wurde, werden zunächst aber noch im nächsten Abschnitt genauer beschrieben.

3.2.6 Statistische Verfahren

Zur Prüfung der Dimensionalität wurde, wie in der Arbeit von Goldston et al. (2013), eine Faktorenanalyse durchgeführt. Die faktorielle Validierung des Ratinginstruments erfolgte dabei auf der Basis der Daten der Teilstichprobe ($n = 70$), bei der das fachdidaktische Wissen zumindest ansatzweise in Unterrichtsplanungen, -materialien und videografierten Planungsgesprächen beobachtet werden konnte.

In einem ersten Schritt wurden die Voraussetzungen der erhobenen Daten in SPSS (IBM Corp., 2016) getestet. Dabei wurden zuerst mittels deskriptiver Statistik die Schiefe und Kurtosis für jeden Indikator bestimmt, zwei Masse, die zeigen, wie stark die Testwertverteilung von der Normalverteilung abweicht (Hartig et al., 2012, S. 93). Gemäss George und Mallery (2010) werden in SPSS für die Prüfung auf Normalverteilung Richtwerte zwischen -2 and +2 als akzeptabel eingestuft. Andere Autoren schlagen Referenzgrössen mit etwas grösseren Werten vor (e.g. Curran, West & Finch, 1996; Kline, 2016). Die vorliegende Studie stützt sich jedoch auf die strengerer Richtwerte gemäss George und Mallery (2010), da die beiden Masse für alle Indikatoren in SPSS ermittelt worden sind. Liegen die Kennwerte zwischen -2 und +2, kann entsprechend von annähernd normalverteilten Daten ausgegangen werden. Anschliessend wurde das Kaiser-Meyer-Olkin-Kriterium (KMO) berechnet, der Bartlett-Test auf Sphärität durchgeführt und die MSA-Koeffizienten (*measure of sampling adequacy*) ermittelt (Bühner, 2011, S. 346 & 347; cf. Nitz, Nerdel & Prechtel, 2012). Der KMO-Wert ist ein Mass für das Korrelationsmuster, das den Daten zugrundeliegt. Es besitzt einen Wertebereich zwischen 0 und 1. Je näher der Koeffizient bei 1 liegt, desto kompakter ist die Korrelationsmatrix und desto grösser auch die Wahrscheinlichkeit, dass distinkte, reliable Faktoren extrahiert werden können. Werte über .80 sind wünschenswert und werden im Bereich .80 - .89 als *meritorious* oder gut, im Bereich $\geq .90$ als *marvellous* oder sehr gut eingestuft. Werte < 0.5 sind für eine Faktorenanalyse inakzeptabel (Backhaus, Erichson, Plinke & Weiber, 2018; Bühner, 2011; Field, 2013). Mit dem

Bartlett-Test auf Sphärität wird die Nullhypothese geprüft, dass alle Korrelationen der Korrelationsmatrix den Wert Null besitzen (Bühl, 2014, S. 628). Ist der Test signifikant, kann diese Hypothese widerlegt und eine Faktorenanalyse durchgeführt werden. Die Berechnung der MSA-Koeffizienten ermöglicht inhaltlich ähnliche Aussagen wie die Bestimmung des KMO-Koeffizienten. Es geht dabei aber nicht um die Frage, ob sich die Korrelationsmatrix für die Faktorenanalyse eignet, sondern es geht um die Eignung jedes einzelnen Indikators für eine Faktorenanalyse. Die Bewertung der MSA-Koeffizienten ist gleich wie für die KMO-Koeffizienten. Werte $< .50$ müssen gemäss Bühner (2011) geprüft, aber nicht zwingend aus dem Test entfernt werden, da es verschiedene Gründe für niedrige Koeffizienten geben kann.

In einem nächsten Schritt wurden die Daten empirisch auf die a priori festgelegte Zuordnung der aus der Literatur hergeleiteten Indikatoren zu Faktoren (Dimensionen) untersucht und damit hinsichtlich ihrer faktoriellen Validität geprüft (Hartig et al., 2012). Dabei wurden in Anlehnung an Asparouhov & Muthén (2009) die Beziehungen (Korrelationen) zwischen den Indikatorvariablen und dem hypothetischen Konstrukt, d.h. den latenten Variablen, mittels *exploratory structural equation modeling* (ESEM) im Programm MPlus Version 8.0 (Muthén & Muthén, 1998-2017), ermittelt. Das ESEM ist eine strukturprüfendes, multivariates statistisches Analyseverfahren, das neben oder an Stelle einer konfirmatorischen Faktorenanalyse auch das Messmodell der explorativen Faktorenanalyse mit Rotationen berücksichtigt (Asparouhov & Muthén, 2009, S. 398). Dabei wird zunächst die Einfachstruktur bzw. die Eindimensionalität des Modells geschätzt und anschliessend die Einfach- bzw. Mehrfachstruktur des Modells durch Rotation in Sinne einer explorativen Faktorenanalyse (EFA) geprüft (Marsh, Morin, Parker & Kaur, 2014). Im zweiten Schritt dieses Verfahrens werden demzufolge auch Nebenladungen zugelassen. Das bedeutet, dass Indikatoren auch auf zwei oder mehrere Faktoren laden können. Das ESEM entspricht damit einer Strukturgleichungsanalyse, die den Zugang zu wichtigen Parametern von Strukturgleichungsmodellen wie Korrelationen der Residuen oder Regressionen zwischen den Faktoren ermöglicht (Asparouhov und Muthén, 2009) und durch den Einbezug der EFA mit Rotation auch gemeinsame Aspekte der jeweils untersuchten Faktoren zulässt. Mit dem integrierten Verfahren des ESEM wird dem in der Literatur viel diskutierten Aspekt Rechnung getragen, dass konfirmatorische Faktorenanalysen zu restriktiv seien, weil sie keine Nebenladungen zulassen (Marsh et al., 2014).

In der vorliegenden Untersuchung wurde das ESEM mit Geomin Rotation in MPlus Version 8.0 (Muthén & Muthén, 1998-2017) durchgeführt. Dabei wurden die Modell-Fits über den Chi-Quadrat-Test, ausgewählte Fit-Indizes des Modells sowie Signifikanztests für die zu schätzenden Modellparameter bestimmt (Bühner, 2011, S. 419). Der Chi-Quadrat-Test

testet die Nullhypothese, dass die Kovarianzmatrix in der Population der vom Modell implizierten Kovarianzmatrix entspricht (Geiser, 2011, S. 60). Erweist sich der Testwert dieser Statistik (exakter Modell-Fit) als signifikant, so bedeutet dies, dass die Abweichungen zwischen impliziter und beobachteter Kovarianzmatrix nicht nur zufallsbedingt sind (Bühner, 2011, S. 407) und dass die Nullhypothese abgelehnt werden sollte. Der Stellenwert dieser Abweichungen sollte daher mit Hilfe anderer Fit-Werte überprüft werden. In der vorliegenden Studie wurden als weitere Kennwerte zur Beurteilung der Modellgüte der CFI (*comparative fit index*), der RMSEA (*root mean square error of approximation*) und der WRMR (*weighted root mean square residual*) berechnet. Der CFI-Wert orientiert sich an einem restriktiveren Nullmodell, in dem alle Variablen als unkorreliert angenommen werden (Bühner, 2011). Er gilt bei komplexen Modellen als sensitiv gegenüber Fehlerspezifikationen und gering sensitiv gegenüber Verteilungsverletzungen und Stichprobengrösse. Der CFI-Wert liegt im Bereich zwischen 0 und 1, sollte aber möglichst nahe bei 1 liegen. Für ein gutes Modell sollte der CFI über .95, besser noch über .97 liegen (Geiser, 2011, S. 60; Bühner, 2011). In der Literatur werden auch Werte $\geq .90$ als gute Modell-Fit-Werte eingestuft (Weiber & Mülhhaus, 2014). Der RMSEA-Wert orientiert sich im Gegensatz dazu an einem saturierten Modell, das die Stichprobenvarianz genau nachbildet und das Modell auf der Grundlage der Daten perfekt beschreibt. Er gibt die durchschnittliche Abweichung pro Freiheitsgrad an und sollte für einen guten Modell-Fit möglichst klein sein (Geiser, 2011). Werte $\leq .050$ gelten als exzellent, Werte zwischen .60-.80 als adäquat, wobei .80 insbesondere bei Stichproben ≤ 250 gilt (Bühner, 2011, S. 425). Für kategoriale Schätzungen (WLS (Weighted Least Squares) und WLSMV (Weighted Least Squares Mean and Variance Adjusted)) ist zudem der WRMR-Wert (Weighted Root Mean Square Residual) entwickelt worden. Dieser sollte idealerweise bei $< .90$ liegen (Urban & Mayerl, 2014).

Neben der Ermittlung der Modell-Fit-Werte wurden auf der Basis der Daten die Faktorladungen berechnet, die angeben, wie gut ein Indikator mit dem Faktor korreliert und diesen repräsentiert (Döring & Bortz, 2016). Die standardisierten Faktorladungen liegen im Wertebereich von -1 bis +1. Eine signifikante Faktorladung von .30 deutet darauf hin, dass eine lineare Beziehung zwischen dem Indikator und dem Faktor besteht. Diese ist grösser Null, kann allerdings die Varianz in der Indikatorvariablen nur gering ($r^2 = 9\%$) erklären. Deshalb werden in der Literatur Richtwerte empfohlen, die möglichst über .70 ($r^2 = 50\%$) liegen, Werte $> .50$ ($r^2 = 25\%$) werden als hinreichend eingestuft (Backhaus et al., 2018; Weiber & Mülhhaus, 2014). Bühner (2011, S. 350) empfiehlt, vor der Eliminierung von Indikatoren mit geringen Faktorladungen ($< .30$) auf die inhaltliche Validität zu achten. Als Mindesteinschätzung für die Reliabilität eines Indikators kann ausserdem gemäss Bühner

(2011) auch die Kommunalität berechnet werden. Sie entspricht dem Anteil der gesamten Varianz einer Variablen und gibt an, in welchem Ausmass eine Variable durch die ermittelten Faktoren erklärt wird. Gerade bei kleinen Stichprobengrössen von $n = 60$ sollten die Kommunalitäten der Indikatoren $h^2 > .60$ sein, denn die Stabilität der Faktorlösung nimmt generell mit wachsender Stichprobengrösse zu (Bühner, 2011, S. 345).

Die Ermittlung der Modellprüfung ist wie bereits erwähnt im Programm MPlus Version 8.0 (Muthén & Muthén, 1998-2017) durchgeführt worden. Dieses Programm erweist sich als flexibel im Umgang mit Datensituationen, welche die strengen Annahmen konventioneller Schätzmethoden wie beispielsweise der multivariaten Normalverteilung der Variablen verletzen (Geiser, 2011, S. 7). In vielen Fällen, bei denen einzelne Items lediglich mit ordinalem Skalenniveau als Indikatoren für latente Faktoren verwendet werden, wie in der vorliegenden Studie, ist die Normalverteilungsannahme nicht haltbar (Finney & DiStefano, 2006). Ratingskalen mit fünf oder weniger Kategorien sind nämlich oft asymmetrisch. Es wurde aus diesem Grund auch bei den Berechnungen in MPlus der robuste Schätzer WLSMV (Weighted Least Squares Mean and Variance Adjusted) verwendet (Kline, 2016; cf. Kotzebue & Nerdel, 2015).

Die Ergebnisse der strukturellen bzw. faktoriellen Validierung werden im Kapitel 4 entlang der Forschungsanliegen 1a und 1b und den dazu formulierten Forschungsfragen 1 und 2 vorgestellt (cf. Kap. I, 2.11).

4 Ergebnisse

Das im Projekt KUBeX entwickelte Ratingmanual zur Erfassung des fachdidaktischen Wissens angehender Lehrpersonen zur Planung von Unterricht zum experimentellen Handeln wurde im vorangehenden Kapitel in Bezug auf die zwei zentralen Gütekriterien, Objektivität und Reliabilität, geprüft. Aufgrund unabhängig erfolgter Ratings an einer kleinen Stichprobe ($n = 23$) von Unterrichtsplanungen, -materialien und Planungsgesprächen konnte die Auswertungsobjektivität bestätigt werden. Ebenso konnte auf der Basis der hoch inferent erhobenen Daten der Teilstichprobe ($n = 70$) eine gute Reliabilität für das untersuchte, latente Konstrukt ermittelt werden (cf. Kap. I, 3.2.5). Das bedeutet, dass die auf der Grundlage des Ratingmanuals erhobenen Daten hinsichtlich der qualitativen Umsetzung des erfassten Konstrukts weiter analysiert und ausgewertet werden können. Das Ratingmanual wurde daher auf seine inhaltliche und kognitive Validität geprüft. Die inhaltliche Validität konnte aufgrund der beschriebenen Vorgehensweise und der damit verbundenen Konzeptspezifikationen bei der Entwicklung des Ratingmanuals als erfüllt eingestuft werden (cf. Kap. I, 3.2.4). Demgegenüber liess sich die kognitive Validität nicht vollumfänglich bestätigen. Der Planungsauftrag, den die angehenden Lehrpersonen auszuführen hatten, hat zwar als diagnostische Aufgabe alle vier Qualitätskriterien nach Shavelson (2010) erfüllt (cf. Kap. I, 3.2.5). Jedoch basierte er auf einer fiktiven Planungssituation (Vignette, cf. Anhang A-1, Abb. A1.2). Entsprechend konnten diese Planungen nicht in unterrichtliches Handeln übersetzt werden. Auch musste der Planungsauftrag zusätzlich zu anderen Aufträgen in den jeweiligen Modulen ausgeführt werden. Es wäre deshalb vor diesem Hintergrund noch zu prüfen gewesen, ob die Qualität der Planungen nicht durch den beschriebenen Kontext beeinflusst worden ist.

Die strukturelle bzw. faktorielle Validität des Instruments bildet den Kern der Teilstudie I. Die folgenden Kapitel befassen sich mit dem Forschungsanliegen 1a und den in diesem Zusammenhang erfolgten Untersuchungen. In Kapitel I, 4.1 werden die Ergebnisse der Validierung des zu prüfenden Strukturmodells berichtet. In einem ersten Schritt werden die Analyseergebnisse der vorangestellten Voraussetzungsprüfung der Daten für eine mögliche Faktorenanalyse vorgestellt (Kap. I, 4.1.1). In Kapitel I, 4.1.2 werden dann die Ergebnisse des ESEM zur Eindimensionalität des Konstruktes dargelegt (Forschungsfrage 1). Kapitel I, 4.1.3 befasst sich mit dem Forschungsanliegen 1b. Dabei werden die Erkenntnisse aus dem ESEM zum 2-Faktorenmodell vorgestellt (Forschungsfrage 2), um dann in

Kapitel I, 4.1.4 die Ergebnisse beider Modelle zu vergleichen. Die zentralen Ergebnisse zu den beiden Forschungsanliegen werden zum Schluss in Kapitel I, 4.2 zusammengefasst.

4.1 Validierung des Strukturmodells zur Erfassung fachdidaktischen Wissens in Unterrichtsplanungen, -materialien und videografierten Planungsgesprächen

Mit den bisherigen Analysen zum Ratingmanual konnte das Strukturmodell und Erhebungsinstrument zwar theoretisch begründet und inhaltlich validiert werden. Es fehlen aber empirische Evidenzen, die bestätigen, dass dieses heuristische Konstrukt auch tatsächlich das entsprechende fachdidaktische Wissen misst. Ziel der Teilstudie I ist daher die strukturelle bzw. faktorielle Validierung dieses Modells. Im Fokus der Forschungsfrage 1 steht zunächst die Prüfung der Validität des gesamten Instruments, das sich aus insgesamt 17 Indikatoren zusammensetzt. Ob diese Indikatoren eine eindimensionale Gesamtskala abbilden und damit das latente Konstrukt messen, soll mit Ergebnissen aus der mittels ESEM erfolgten Untersuchung der Eindimensionalität des Konstruktes beantwortet werden. Die Beantwortung der Forschungsfrage 2 nämlich, ob sich die eindimensionale Gesamtskala auch sinnvoll in Subskalen des theoretischen Konstrukts differenzieren lässt und sich dabei die beiden Dimensionen des Planungsmodells der «Didaktischen Rekonstruktion», «Analyse» und «Konstruktion», aus den Daten extrahieren lassen, soll ebenfalls mittels Ergebnissen aus dem ESEM erfolgen. Beide Untersuchungen sind, wie in Kap. I, 3.2.6 erwähnt, an der Teilstichprobe $n = 70$ erfolgt. Alle Ergebnisse werden entlang der beiden Fragestellungen vorgestellt und erläutert (cf. Kap. I, 2.11). Zunächst werden aber noch die Ergebnisse der Analysen berichtet, mit deren Hilfe die Eignung der Daten für die faktorielle Validierung geprüft worden ist.

4.1.1 Datenanalyse – Voraussetzung für die faktorielle Validierung

Zur Prüfung der faktoriellen Validität des Ratingmanuals wurden die Daten auf deren Eignung analysiert. Die Berechnung der Schiefe und Kurtosis aller Indikatoren hat ergeben, dass lediglich die Werte der beiden Indikatoren «Kenntnisse» der Kategorie «Schülervorstellungen» und «Prozessreflexion» der Kategorie «Reflexivität» stark über den Richtwerten lagen. Der Indikator «Kenntnis» wies eine Schiefe von 4.77 ($SE=0.28$) und eine Kurtosis von 24.27 ($SE=0.56$) auf, während die Werte für den Indikator «Prozessreflexion» für die Schiefe bei 2.96 ($SE=0.56$) und für die Kurtosis bei 6.80 ($SE=0.56$) lagen. Diese beiden Indikatoren wurden deshalb für die weiteren Analysen aus dem Test entfernt. Für die verbleibenden 15 Indikatoren wurde anschliessend mit dem KMO-Kriterium und dem Bartlett-Test auf Sphärität die Eignung der Korrelationsmatrix ermittelt, die den Daten zu-

grundliegt. Beide statistischen Prüfverfahren haben gezeigt, dass die Daten für die Durchführung einer Faktorenanalyse geeignet sind ($KMO = 0.89$, Bartlett-Test: $X^2(105, n = 70) = 779.85, p < .001$). Die Analyse der MSA-Werte (*measure of sampling adequacy*) hat auch ergeben, dass die Werte von 5 Indikatoren $> .90$, 8 Werte $> .80$ und 2 Werte $> .70$ sind (cf. Anhang A-1, Tab. A2.1). Die 15 Indikatoren können damit für die Faktorenanalyse verwendet werden (Kap. I, 3.2.6). Die in Kap. 3.2.5.2 berichteten Trennschärfen stützen diesen Befund. Sie liegen für diese 15 Indikatoren alle $> .30$.

Mit den 15 Indikatoren, welche die latente Variable «Fachdidaktisches Wissen zur Planung von Unterricht zum experimentellen Handeln» erfassen sollen, wurde ein ESEM mit Geomin-Rotation durchgeführt. Das ESEM erfolgte im Programm MPlus Version 8.0 (Muthén & Muthén, 1998-2017) und konnte ohne fehlende Werte realisiert werden. Die Ergebnisse zur ersten Forschungsfrage 1 werden im nächsten Abschnitt vorgestellt.

4.1.2 Faktorielle Validierung des Ratingmanuals (1-Dimensionalität)

Die Ergebnisse zur Forschungsfrage 1, ob sich das fachdidaktische Wissen zur Planung einer Lerngelegenheit bzw. Unterrichtsstunde zum experimentellen Handeln mit Hilfe des im Projekt KUBeX entwickelten hoch inferenten Instruments in Unterrichtsplanungen auf der Grundlage der Daten valide erfassen lässt, sind in Tabelle 11 dargestellt.

Tabelle 11: Modell-Fit-Statistik 1-Faktormodell

	1-Faktormodell
X^2	116.730*
df	90.00
X^2/df	1.297
$p(X^2)$.031
RMSEA (90% Konfidenzintervall)	.065
$p(RMSEA \leq 0.05)$.233
CFI	.993
WRMR	.712

Aus der Modell-Fit-Statistik lässt sich ableiten, dass das 1-Faktormodell mehrheitlich auf die hoch inferent erhobenen Daten passt. Wenn auch der Chi-Quadrat-Wert für das Zielmodell ($X^2 = 116.73, df = 90, p = .031$) noch darauf hinweist, dass das Modell knapp signifikant von den Daten abweicht, so vermitteln die übrigen Fit-Werte ein anderes Bild. Diese weisen auf einen guten Modell-Fit hin. Der CFI ist mit $.993 \geq .95$, was einem exzellenten Modell-Fit entspricht. Die Annahme, dass das Nullmodell zutrifft und keine

korrelativen Zusammenhänge zwischen den Variablen bestehen, ist damit nicht haltbar. Der RMSEA liegt bei .065, d.h. er liegt für eine Stichprobengröße von $n = 70$ im adäquaten Bereich. Die Nullhypothese muss damit nicht verworfen werden, da die Wahrscheinlichkeit bei $p = .233$ liegt. Der WRMR liegt mit < 1 ebenfalls im akzeptablen Wertebereich.

Diese Befunde werden auch durch die im ESEM ermittelten Faktorladungen gestützt. Sie sind in Abbildung 21 dargestellt.

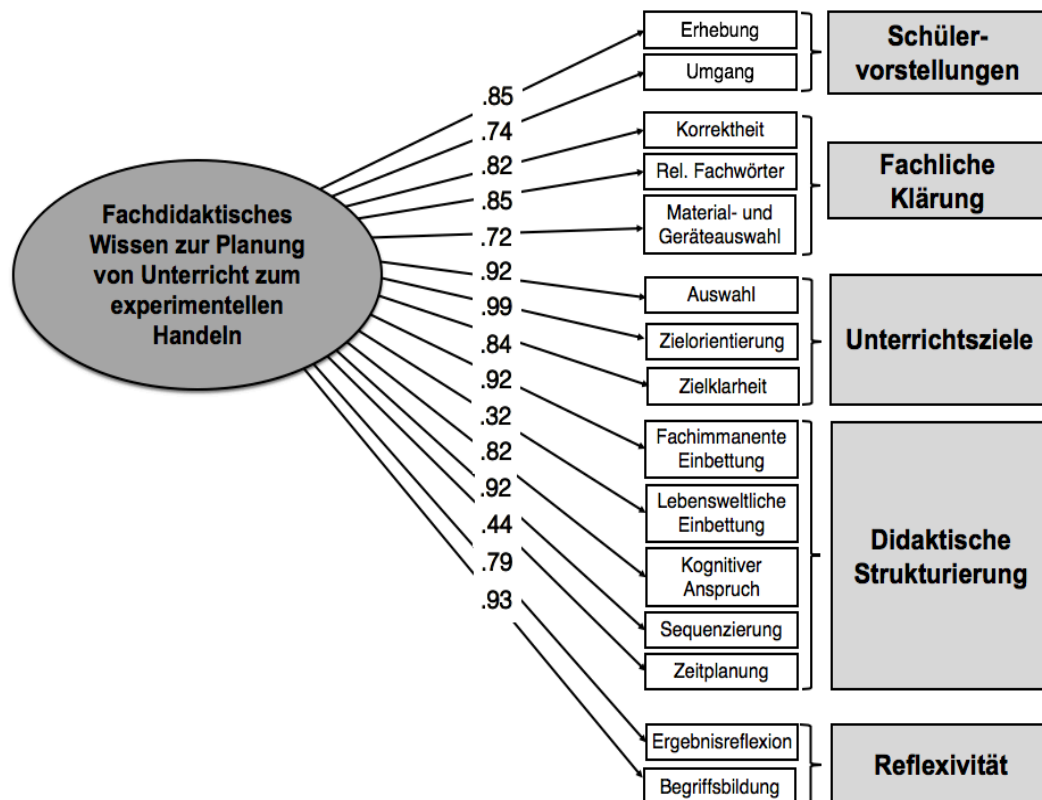


Abbildung 21: 1-Faktormodell «Fachdidaktisches Wissen zur Planung von Unterricht zum experimentellen Handeln» mit Faktorladungen (Reliabilität des Konstruktes Cronbachs $\alpha = .93$)

Die Faktorladungen der 15 Indikatoren bewegen sich im Bereich zwischen .32 und .99 und laden alle auf den untersuchten Faktor, die latente Variable «Fachdidaktisches Wissen zur Planung von Unterricht zum experimentellen Handeln». Die höchsten Faktorladungen liegen über 0.9. Dazu gehören die Korrelationen zu Indikatoren der Kategorien:

- Unterrichtsziele («Auswahl» und «Zielorientierung»)
- Didaktische Strukturierung («Fachimmanente Einbettung» und «Sequenzierung»)
- Reflexivität («Begriffsbildung»).

5 Faktorladungen liegen $> .80$ und 3 $> .70$. Die beiden niedrigsten Faktorladungen sind bei den Indikatoren «Lebensweltliche Einbettung» und «Zeitplanung» zu verzeichnen, beides Indikatoren der Kategorie «Didaktische Strukturierung». Sie liegen bei $.32$ und $.44$. Von den 15 Faktorladungen sind demzufolge 13 Faktorladungen im erwünschten Bereich von $> .70$ angesiedelt (Kap. I, 3.2.6). Die beiden tieferen Faktorladungen werden aufgrund ihrer inhaltlichen Validität nicht aus dem Test entfernt. Die Reliabilität des über die 15 Indikatoren beschriebenen Konstruktes «Fachdidaktisches Wissen zur Planung von Unterricht zum experimentellen Handeln» beträgt $.93$ und ist damit sehr gut.

Zusammenfassend kann festgehalten werden, dass aufgrund der dargelegten Ergebnisse die im Ratinginstrument enthaltenen Indikatoren die latente Variable «Fachdidaktisches Wissen zur Planung von Unterricht zum experimentellen Handeln» abbilden. Das bedeutet, dass sich die Wissensdimension über das hoch inferente Ratingmanual operationalisieren und sich im Rahmen der Performanz messen lässt. Das im Projekt KUBeX entwickelte Ratingmanual ist damit faktoriell valide und erfüllt auf der Grundlage der Teilstichprobe ($n = 70$) die meisten der Gütekriterien. Entsprechend kann das Ratingmanual zur Erfassung der Qualität «Fachdidaktischen Wissens zur Planung von Unterricht zum experimentellen Handeln» in Unterrichtsplanungen, Arbeitsmaterialien und videografierten Planungsgesprächen eingesetzt werden. Der hohe Cronbachs α -Wert von $.93$ kann damit auch einer hohen Messgenauigkeit des Instruments zugeschrieben werden. Allerdings ist die interne Konsistenz kein Mass dafür, ob die Skala tatsächlich nur ein Merkmal erfasst. Bei der Interpretation der Höhe von Reliabilitätskoeffizienten müssen daher immer auch die Beschaffenheit der Kategorien sowie die Anzahl der Merkmalsausprägungen geprüft werden (Kap. I, 3.2.5). Die Ergebnisse zur Frage der Dimensionalität des Strukturmodells und damit auch zur Forschungsfrage 2 werden vor diesem Hintergrund im nächsten Abschnitt vorgestellt.

4.1.3 Faktorielle Validierung des Ratingmanuals (2-Dimensionalität)

Die Ergebnisse zur Forschungsfrage 2, inwieweit die Daten aus der Beurteilung der Unterrichtsplanungen, -materialien und videografierten Planungsgespräche angehender Lehrpersonen die zentralen Wissensdimensionen «Analyse» und «Konstruktion» abbilden, die im ERTE-Modell enthalten sind, sind in Tabelle 12 dargelegt.

Die Modell-Fit-Statistik zeigt, dass das 2-Faktorenmodell ($X^2 = 86.33$, $df = 76$, $p = .196$) noch besser auf die zugrundeliegenden, hoch inferent erhobenen Daten passt. Der Chi-Quadrat-Test ist nicht signifikant, das bedeutet, dass die Abweichungen zwischen impliziter und beobachteter Kovarianzmatrix nur zufallsbedingt sind. Der CFI-Wert, der $> .95$ bzw. $> .97$ sein sollte, ist mit $.997$ exzellent. Ausserdem liegt der RMSEA deutlich

unter .05. Die Nullhypothese $RMSEA \leq .05$ muss deshalb nicht verworfen werden, da die Wahrscheinlichkeit bei $p = .562$ liegt. Der WRMR liegt mit .536 im akzeptablen Bereich.

Tabelle 12: Modell-Fit-Statistik 2-Faktorenmodell

	2-Faktorenmodell
χ^2	86.334
<i>df</i>	76.00
χ^2/df	1.136
$p(\chi^2)$.196
RMSEA (90% Konfidenzintervall)	.044
$p(RMSEA \leq 0.05)$.562
CFI	.997
WRMR	.536

Aufgrund der Ergebnisse kann abgeleitet werden, dass sich die beiden, im ERTE-Modell enthaltenen Wissensdimensionen bzw. Skalen, «Analyse» und «Konstruktion», auf der Basis der erhobenen Daten extrahieren lassen. Der Faktor «Analyse» wird durch 9 Indikatoren aus fünf Kategorien, der Faktor «Konstruktion» durch 5 Indikatoren aus drei Kategorien abgebildet (Abb. 22). Damit laden insgesamt 14 von 15 untersuchten Indikatoren signifikant auf einen der beiden Faktoren. Eine Ausnahme bildet der Indikator «Zeitplanung», der für keinen der beiden Faktoren eine signifikante Faktorladung aufweist und deshalb nicht mehr in der Abbildung 22 aufgeführt ist.

Aus den Korrelationen der 14 Indikatoren mit den Faktoren wird ersichtlich, dass die Indikatoren der Kategorien «Schülervorstellungen» und «Reflexivität» ausschliesslich auf den Faktor «Analyse» laden und Faktorladungen im Bereich zwischen .55 und .96 aufweisen. Demgegenüber korrelieren die Indikatoren der anderen drei Kategorien «Fachliche Klärung», «Unterrichtsziele» und «Didaktische Strukturierung» jeweils mit einem der beiden Faktoren und dies mit Faktorladungen, die im Bereich zwischen .53 und .91 liegen. Insgesamt 10 Indikatoren weisen Faktorladungen über .70 und nur 4 Indikatoren Faktorladungen über .50 auf (Tab. 13). Damit sind alle Faktorladungen hinreichend bis sehr gut. Die aufgrund der Faktorladungen erfolgte Zuteilung der einzelnen Indikatoren zu den beiden Faktoren ist inhaltlich mit den beiden im Modell der «Didaktischen Rekonstruktion» bedeutsamen Wissensdimensionen «Analyse» und «Konstruktion» vereinbar. Indikatoren aller 5 Kategorien, die mit den Kompetenzen der Lernenden, den Kompetenzansprüchen im Fach und den Vermittlungsabsichten korrelieren, bilden den Faktor «Analyse» ab.

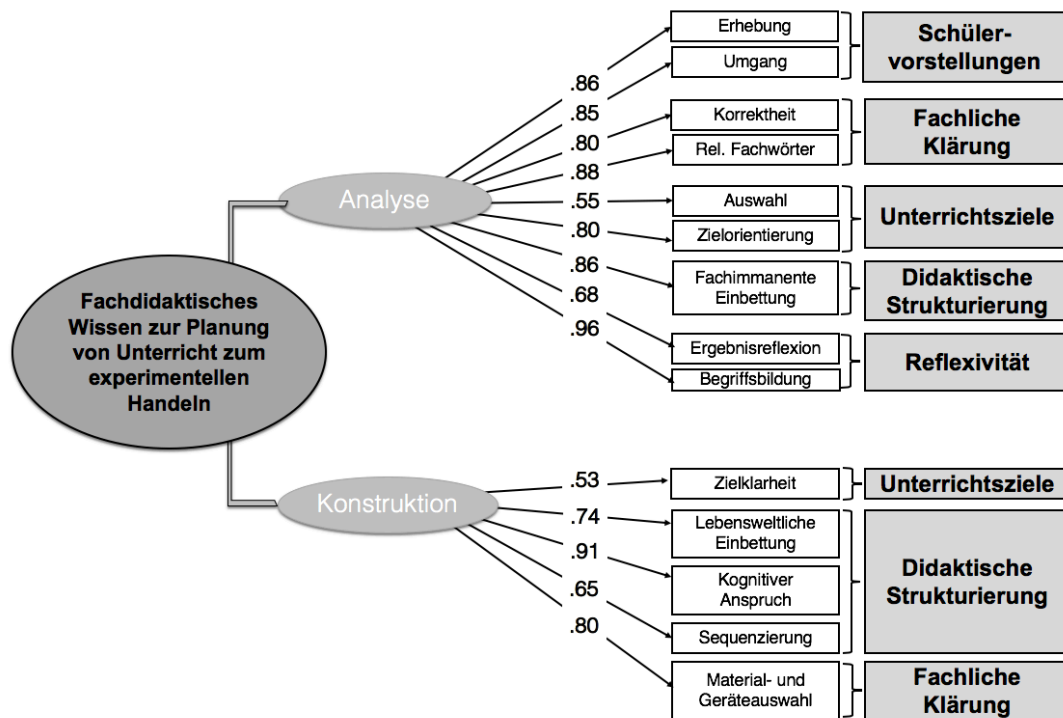


Abbildung 22: 2-Faktorenmodell «Fachdidaktisches Wissen zur Planung von Unterricht zum experimentellen Handeln» mit Faktorladungen (Reliabilität der Faktoren «Analyse» (Cronbachs $\alpha = .94$), «Konstruktion» (Cronbachs $\alpha = .83$))

Indikatoren, die eher in Zusammenhang mit der didaktischen Strukturierung der Lernprozesse stehen, korrelieren mit dem Faktor «Konstruktion». Der Faktor «Analyse» erklärt dabei insgesamt 59.4% der Gesamtvarianz, während mit dem Faktor «Konstruktion» 27.2% der gesamten Varianz aufgeklärt wird. Bei genauerer Betrachtung der Faktorwerte fällt allerdings auf, dass die beiden Indikatoren «Auswahl» der Kategorie «Unterrichtsziele» und «Sequenzierung» der Kategorie «Didaktische Strukturierung» signifikant auf beide ermittelten Faktoren laden. Während ersterer mit einer höheren Faktorladung auf den Faktor «Analyse» lädt, korreliert letzterer aufgrund der höheren Faktorladung stärker mit dem Faktor «Konstruktion». Der Entscheid für die Zuordnung ist in Anlehnung an die Richtwerte erfolgt (Backhausen et al., 2018, Weiber et al., 2014). Anzumerken ist auch, dass bei 10 von 14 Indikatoren Kommunalitäten von $> .60$ vorliegen. Nur die Indikatoren «Auswahl der Ziele», «Ergebnisreflexion», «Zielklarheit» und «Sequenzierung» zeichnen sich durch etwas tiefere Kommunalitätswerte aus. Diese Indikatoren laden mit Ausnahme des Indikators «Ergebnisreflexion» auf beide Faktoren vergleichsweise hoch. Der Indikator «Zeitplanung» weist, was nicht erstaunt, einen sehr tiefen Kommunalitätswert auf.

Tabelle 1: 2-Faktorenmodell – Faktorladungen und Kommunalitäten der Indikatoren

Faktor / Skala	Indikator	Faktorladungen		h^2
		1	2	
Analyse	Erhebung Schülervorstellungen	.86*	.01	.75
	Umgang mit Schülervorstellungen	.85*	-.11	.73
	Korrektheit	.80*	.06	.64
	Relevante Fachwörter	.88*	-.01	.78
	Auswahl Ziele	.55*	.43*	.49
	Zielorientierung	.80*	.25	.70
	Fachimmanente Einbettung	.86*	.09	.75
	Ergebnisreflexion	.68*	.14	.48
Konstruktion	Begriffsbildung	.96*	-.02	.92
	Zielklarheit	.37	.53*	.42
	Lebensweltliche Einbettung	-.35	.74*	.67
	Kognitiver Anspruch	.01	.91*	.83
	Sequenzierung	.35*	.65*	.55
	Material- und Geräteauswahl	.00	.80*	.64
Zeit		.10	.39	.16
Aufgeklärte Varianz in %		59.4	27.2	

Anmerkung: Faktorladungen > .50 sind durch Fettdruck hervorgehoben. * = signifikante Ladung

Zusammenfassend lässt sich festhalten, dass die Analyse des 2-Faktorenmodells ebenfalls sehr gute Fit-Werte zeigt. Es lassen sich aus den erhobenen hoch inferenten Daten zwei Dimensionen extrahieren. Das latente Merkmal «Fachdidaktisches Wissen zur Planung von Unterricht zum experimentellen Handeln» kann demzufolge in zwei Subkategorien aufgliedert werden, die beide der Tiefenstruktur des Unterrichts zugeordnet werden können. Die Dimension der «Analyse» wird durch Indikatoren beschrieben, die für die Anlage eines Unterrichts bzw. einer Lerngelegenheit von Bedeutung sind. Im Zentrum der Dimension der «Konstruktion» stehen demgegenüber Merkmale, die hauptsächlich für den Prozess der didaktischen Strukturierung des Unterrichts bzw. der Lerngelegenheit wichtig sind. Beide Faktoren weisen gute Reliabilitäten aus. Diese liegen für den Faktor «Analyse» bei .94 und für den Faktor «Konstruktion» bei .83.

Die Frage, ob sich die beiden Modelle auch statistisch unterscheiden, ist im Programm MPlus Version 8.0 (Muthén & Muthén, 1998-2017) geprüft worden. Die Ergebnisse werden im nächsten Abschnitt präsentiert.

4.1.4 1-Faktor- und 2-Faktorenmodell im Vergleich

Die im Programm MPlus geschätzten Fit-Indizes sind in Tabelle 14 einander gegenübergestellt.

Tabelle 14: Goodness-of-Fit-Indizes der beiden Modelle ($n = 70$)

Modell	df	χ^2	p
1-faktoriell	90	116.730*	.031
2-faktoriell	76	86.334	.196

Die Goodness-of-Fit-Indizes der beiden Modelle deuten darauf hin, dass sowohl das 1-Faktormodell, wie auch das 2-Faktorenmodell auf die hoch inferent erhobenen Daten passen, wobei beim 1-Faktormodell der Chi-Quadrat-Wert noch knapp signifikant ist und nur die anderen Fit-Werte überzeugen.

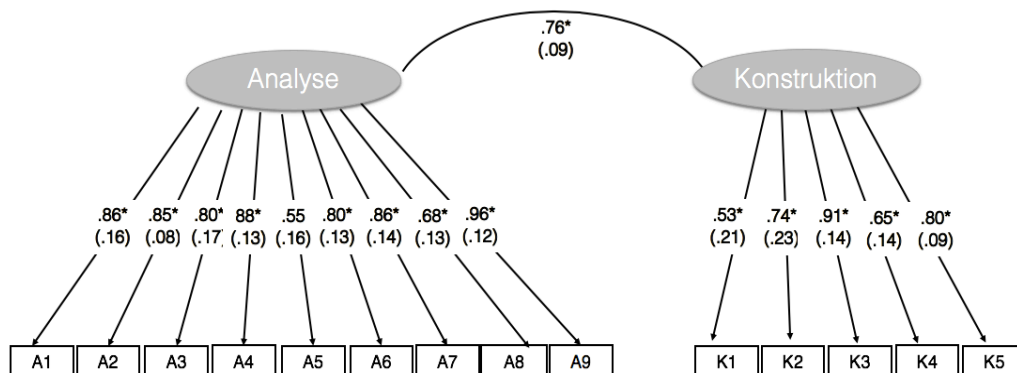


Abbildung 23: 2-Faktorenmodell mit standardisierten Koeffizienten für die Faktoren «Analyse» und «Konstruktion» zur Untersuchung von Unterrichtsplanungen, -materialien und videografierten Planungsgesprächen von angehenden Lehrpersonen ($n = 70$, Ratingmanual mit 14 Indikatoren (A1-A9) für den Faktor «Analyse» und K1-K5 für den Faktor «Konstruktion», Standardfehler sind in Klammern angegeben, * $p \leq .05$)

Der statistische Vergleich beider Modelle zeigt denn auch, dass sich die beiden Modelle signifikant voneinander unterscheiden ($\chi^2 = 27.163$, $df = 14$, $p = .016$). Das 1-Faktormodell weist im Vergleich zum 2-faktoriellen Modell insgesamt eine etwas schlechtere Passung auf.

Das in Abbildung 23 mit den standardisierten Koeffizienten abgebildete 2-faktorielle Modell zum «Fachdidaktischen Wissen zur Planung von Unterricht zum experimentellen

Handeln» wird daher, trotz der relativ hohen Korrelation von .76 ($SE = .09$) zwischen den beiden Faktoren, präferiert.

4.2 Zusammenfassung der Ergebnisse der Teilstudie I

Ziel der in Kapitel 4 dokumentierten faktoriellen Validierung war die explorative Prüfung des im Projekt KUBeX entwickelten Ratingmanuals zur Erfassung von fachdidaktischem Wissen in Unterrichtsplanungen, -materialien und videografierten Planungsgesprächen angehender Lehrpersonen. Die Ergebnisse zur Forschungsfrage 1 zeigen, dass sich auf der Basis der erhobenen Daten das fachdidaktische Wissen zur Planung von Unterricht zum experimentellen Handeln als latente Variable mit Hilfe des Ratingmanuals als 1-faktorielles Konstrukt valide erfassen lässt. Die Goodness-of-Fit-Werte sind dabei mehrheitlich im akzeptablen Bereich. Die Ergebnisse zur Forschungsfrage 2, mit der die Beschaffenheit der Kategorien sowie die Anzahl der Merkmalsausprägungen geprüft wurde, deuten weiter darauf hin, dass auch ein 2-faktorielles Modell auf die erhobenen Daten passt. Die Goodness-of-Fit-Werte sind hier vergleichsweise sogar besser, was statistisch durch einen Modellvergleich bestätigt werden konnte. Das 2-Faktorenmodell unterscheidet sich signifikant vom 1-Faktormodell, d.h. es lassen sich demzufolge zwei Faktoren aus den erhobenen Daten extrahieren. Dabei ist die im 2-Faktorenmodell auf der Basis der Faktorladungen erfolgte Zuteilung der Indikatoren zu den Faktoren inhaltlich vereinbar mit den Wissensdimensionen «Analyse» und «Konstruktion» aus dem Modell der «Didaktischen Rekonstruktion» (Kap. I, 2.8.4). Die hier vorgestellten Befunde werden im nächsten Kapitel vor dem Hintergrund des aktuellen Forschungsstandes näher beleuchtet und diskutiert.

5 Diskussion

Das Planungshandeln im Sinne eines antizipierenden Probehandelns erfordert die Ausdifferenzierung, Integration und Transformation des Professionswissens, insbesondere der drei Wissensdimensionen nach Shulman (1987) (Seel, 2011). Für die Ausübung der Lehrprofession wird dabei das fachdidaktische Wissen als wichtige Grundlage eingestuft (Kind, 2009). Stender et al. (2017) konnten zeigen, dass Unterrichtsplanungen als Prozess für die Interpretation und Kompilierung des Professionswissens, insbesondere des themenspezifischen fachdidaktischen Professionswissens, angesehen werden können. Auch gemäss Baer et al. lässt sich Wissen auf der Basis von Kriterien qualitativ hochwertigen Unterrichts in Unterrichtsplanungen erfassen (Baer et al., 2011). Damit geben Unterrichtsplanungen Einblick in die Kompetenzen und das handlungsvorbereitende Professionswissen der Studierenden (cf. Terbrügge, 2001; Tillema, 2009). Die Kompetenz des Planungshandelns lässt sich demzufolge an der Qualität der gestalteten Handlungspläne ableiten. Entsprechend stellen Unterrichtsplanungen ein zentrales Element auf dem Weg der Professionalisierung zum Lehrerberuf dar (John, 2006; Mutton, Hagger & Burn, 2011). Für die Entwicklung fachdidaktischen Wissens, das als wichtige Grundlage für die Ausübung der Lehrprofession angesehen wird (Kind, 2009), kommt dem Planungshandeln daher eine wesentliche Bedeutung zu (Ball & Bass, 2000; Baylor, 2002). Trotz dieser anerkannten Bedeutung erfährt das Planungshandeln von (angehenden) Lehrpersonen wie auch die Entwicklung der Fähigkeit zur Unterrichtsplanung in der Forschung vergleichsweise wenig Aufmerksamkeit. Beschreibungen der dafür erforderlichen Expertise, insbesondere zum Planen von Unterricht zum experimentellen Handeln als Fachinhalt sind derzeit rar. Demgemäss sind geeignete Erhebungsinstrumente, mit deren Hilfe eine standardisierte Erfassung der Qualität der Planungsperformanz in Bezug auf das beobachtbare fachdidaktische Wissen zum experimentellen Handeln erfolgt, selten und inkonsistent (Hasse et al., 2014; Vogelsang & Riese, 2017; Weingarten & van Ackeren, 2017). Dies erstaunt, weil die Qualität der in der Lehrpersonenbildung erworbenen professionellen Fähigkeiten und die sich daraus entwickelnden prozeduralen Routinen das unterrichtliche Handeln von Lehrpersonen beeinflussen (Baumert et al., 2011; Pauli & Reusser, 2009). Aktuell geben Studien Anlass zur Annahme, dass sich lernwirksames Lehren und Unterrichten auf das Wissen und Können von Lehrpersonen zurückführen lässt (cf. Besser & Krauss, 2009; Bromme, 2008; Jüttner & Neuhaus, 2013) und dass diese Kompetenzen wiederum Wirkung auf die Leistungen der Schülerinnen und Schüler haben (Baumert & Kunter, 2013b; Lipowsky et al., 2009; cf. auch Weinert et al., 1990).

Vor diesem Hintergrund ist im Projekt KUBeX ein Ratingmanual entwickelt worden, mit dessen Hilfe das fachdidaktische Wissen von angehenden Lehrpersonen zum experimentellen Handeln in Unterrichtsplanungen hoch inferent erfasst werden kann (Kap. I, 3.1). Ziel der Teilstudie I ist eine erste empirische Prüfung dieses neuen Ratingmanuals, das grösstenteils auf der Basis des Planungs- und Forschungsmodells der «Didaktischen Rekonstruktion» normativ festgelegt und modelliert worden ist. Die in Kapitel I, 4 dargelegten Ergebnisse der durchgeführten explorativen, faktoriellen Validierung dieses Strukturmodells werden unter Einbezug des aktuellen Forschungsstandes in den nächsten Abschnitten diskutiert und eingeordnet.

5.1 Befunde zur faktoriellen Validierung des Ratingmanuals

Für die Erfassung der Planungstätigkeit bzw. Planungskompetenz ist in der Literatur kein einheitliches Konzept vorzufinden (König et al., 2015; cf. Stender, 2014, S. 87f.). Der Grund dafür liegt in den verschiedenen Forschungsanliegen und den damit verbundenen unterschiedlichen Forschungsfragen. Eine Folge davon sind die verschiedenen methodischen Herangehensweisen und Untersuchungsdesigns (cf. Kap. I, 2.8.3). Studien, die den Fokus auf eine standardisierte Analyse konkret ausgearbeiteter Planungsunterlagen legen und zu diesem Zweck ein Kodiermanual mit Indikatoren zur Rekonstruktion und Quantifizierung der Planungsentscheidungen entwickelt haben, sind vergleichsweise selten. Demgemäss gibt es in der Literatur auch nur wenige Studien, die das realisierte Planungshandeln mit ausgewähltem, fachspezifischem Fokus (hier Naturwissenschaften & Technik) analysiert haben (Gess-Newsome et al., 2017; Goldston et al., 2013; Hacker & Sova, 1998; Jacobs et al., 2008; Ozogul, Olina, et al., 2008; van der Valk & Broekman, 1999). Die Planung von Unterricht zum experimentellen Handeln als Fachinhalt ist unter diesen Studien nur in einer Arbeit von Goldston et al. untersucht worden (Goldston et al., 2013). Das Ratingmanual beruhte nicht auf den Wissensdimensionen von Shulman (1987), wurde aber faktoriell validiert. Demgegenüber ist in einer etwas älteren Studie von van der Valk und Broekman (1999) das fachdidaktische Wissen bzw. PCK mit den beiden zentralen Wissensfacetten von Shulman (1987) in erstellten Unterrichtsplanungen sowie in Interviews untersucht worden. Im Fokus stand aber nicht das experimentelle Handeln und das Instrument wurde auch nicht validiert. In einer aktuellen Interventionsstudie von Gess-Newsome et al. (2017) wurde zur Erfassung von PCK von Biologielehrpersonen der Sekundarstufe I ein Ratinginstrument entwickelt.

Dieses Instrument ist mit Bezug zu ausgewählten biologischen Fachinhalten themenspezifisch angelegt worden und besteht aus drei Dimensionen. Von den drei festgelegten Dimensionen des Ratinginstruments konnten mittels Faktorenanalyse insgesamt zwei (PCK-CK und PCK-PK) bestätigt werden.

Die beschriebene Befundlage zu standardisierten Analysen von Unterrichtsplanungen zeigt, dass bislang keine konsistente Beschreibung des fachdidaktischen Wissens zur Planung von Unterricht zum experimentellen Handeln von (angehenden) Lehrpersonen vorliegt. Das im Projekt KUBeX entwickelte Ratinginstrument zur Erfassung und Beurteilung dieser Wissensdimension und die hier erfolgte erste Validierung des Instruments stellt daher ein erster Schritt in diese Richtung dar. Die Ergebnisse der Validierung des Instruments werden im nächsten Abschnitt entlang der beiden Forschungsanliegen 1a und 1b der Teilstudie I und der bereits dargelegten Literatur in Kap. I, 2 diskutiert.

5.1.1 Faktorielle Validierung des Manuals (Forschungsanliegen 1a)

Für die Modellierung des neuen Strukturmodells sind, wie in Kap. I, 3.2.4 berichtet, mehrheitlich die Wissensfacetten herangezogen worden, die für die Planung eines qualitativollen Unterrichts von Bedeutung sind. Das fachdidaktische Wissen wurde in Anlehnung an Modellierungen von PCK in Bezug auf die Wissensfacetten eindimensional konzeptualisiert, da der inhaltliche Fokus mit dem Thema Erkenntnisgewinnung vorgegeben war und Überlegungen zu den Wissenstypen im Modell nicht berücksichtigt worden sind (cf. Kap. I, 2.5.2). Als Vorlage für die Auswahl der Wissensfacetten wurden das Planungs- und Forschungsmodell der «Didaktischen Rekonstruktion» und in diesem Zusammenhang Kriterien, die die Tiefenstruktur des Unterrichts beschreiben, beigezogen (Kattmann et al., 1997). Entsprechend wurden neben den Kernfacetten von Shulman (1987) auch weitere, für den Planungsprozess wichtige Kriterien ausgewählt und operationalisiert (cf. Kap. 2.5.2). Daraus resultierte ein Strukturmodell, mit dessen Hilfe das «Fachdidaktische Wissen zur Planung von Unterricht zum experimentellen Handeln» als latente Variable erfasst werden soll. Die latente Variable wurde initial mit 19 Indikatoren, die theoriebasiert begründet und 5 Kategorien zugeordnet worden sind, im Sinne einer idealtypischen Unterrichtsplanung modelliert. Die Qualität der 19 Indikatoren wurde entlang einer vierstufigen Ratingskala (1 bis 4) ausdifferenziert, die Qualitätsstufen möglichst genau beschrieben. Die inhaltliche Validität des Strukturmodells konnte mit der dargelegten Vorgehensweise und den erfolgten Konzeptspezifikationen bei der Entwicklung des Ratingmanuals bestätigt werden (cf. Kap. I, 3.2.4). Erste Analysen der Daten, die mit Hilfe des eventbasierten, hoch inferenten Ratings erhoben wurden, haben jedoch ergeben, dass in 49 von total 119 Unterrichtsplanungen, -materialien und videografierten Planungsgesprächen keine Hinweise auf

einen Unterricht zu beobachten sind, der im Zeichen des experimentellen Handelns steht. Der sich in dieser Form äussernde Bodeneffekt kann darauf zurückzuführen sein, dass das Ratingmanual nur das «Fachdidaktische Wissen zur Planung von Unterricht zum experimentellen Handeln» erfasst und zu wenig genau zwischen dem Vorhandensein oder Fehlen des experimentellen Handelns als Fachinhalt in Unterrichtsplanungen, -materialien und videografierten Planungsgesprächen unterscheidet. Da der fachinhaltliche Fokus der Teilstudie I auf dem experimentellen Handeln liegt, wurde die grössere Teilstichprobe ($n = 70$), die sich signifikant von der anderen Teilpopulation ($n = 49$) unterscheidet (cf. Kap. II, 3.1.1) und bei der das experimentelle Handeln zumindest in Ansätzen beobachtet werden konnte, als Grundlage für die weiteren Analysen verwendet.

Wie die Ergebnisse der statistischen Berechnungen zur Forschungsfrage 1 zeigen, kann mit Hilfe des Ratingmanuals das «Fachdidaktische Wissen zur Planung von Unterricht zum experimentellen Handeln» valide erfasst werden. Allerdings haben Analysen zum Erhebungsinstrument im Vorfeld der Validierung ergeben, dass sich insgesamt vier Indikatoren nicht für die faktorielle Validierung eignen. Einerseits konnten die Indikatoren «Lehrplanbezug» und «Individualisierung» nicht im Messmodell berücksichtigt werden, weil in den Unterrichtsplanungen, -materialien sowie videografierten Planungsgesprächen dazu keinerlei Angaben gemacht worden sind (Kap. I, 3.2.4). Dies kann möglicherweise darauf zurückzuführen sein, dass die Studierenden ihre Planungen aufgrund einer Vignette erstellen mussten, in der wichtige Eckdaten, die sie für die Planung benötigten, bereits vorgegeben waren. Denn mit dem Thema «Experimentieren - Lernen am Beispiel Akkommodation oder Adaptation» war implizit bereits ein Ziel angedeutet, weshalb der Lehrplanbezug von den Studierenden vermutlich nicht mehr explizit in den Planungen aufgeführt worden ist. Auch scheint die, in der Vignette weiter aufgeführte Angabe, dass die SuS der fiktiven Klasse ein durchschnittliches bis gutes Leistungsniveau aufweisen, die Studierenden nicht veranlasst zu haben, Individualisierungsüberlegungen in den Planungen zu vermerken. Aufgrund der vor dem ESEM im Programm MPlus Version 8.0 (Muthén & Muthén, 1998-2017) weiter erfolgten Eignungsprüfung der Daten, konnten andererseits die beiden Indikatoren «Kenntnis» der Kategorie «Schülervorstellungen» und «Prozessreflexion» der Kategorie «Reflexivität» mit ihren stark über den Richtwerten liegenden Kennwerten «Schiefe» und «Kurtosis» ebenfalls nicht für die faktorielle Validierung berücksichtigt werden. Ein Grund dafür könnte darin zu suchen sein, dass das entsprechende Wissen zu diesen beiden Indikatoren ebenfalls wenig in den Planungen zu beobachten war. Die Ratings dieser beiden Indikatoren waren insgesamt schlecht, was sich in tiefen Mittelwerten äusserte.

Das ESEM ist daher mit 15 Indikatoren auf der Basis der erhobenen Daten und ohne fehlende Werte in MPlus Version 8.0 (Muthén & Muthén, 1998-2017) für die Teilstichprobe $n = 70$ durchgeführt worden. Dabei zeigen die Goodness-of-Fit-Indizes des 1-Faktormodells mehrheitlich gute Fit-Werte. Obgleich der Chi-Quadrat-Test noch knapp signifikant ist, weshalb das Modell im Grunde verworfen werden müsste, weisen die anderen Fit-Indizes sehr gute Werte auf. Das bedeutet, dass die im Ratinginstrument verbleibenden 15 Indikatoren die latente Variable «Fachdidaktisches Wissen zur Planung von Unterricht zum experimentellen Handeln» trotzdem abbilden. Die Wissensdimension lässt sich demnach über das hoch inferente Ratingmanual operationalisieren und im Rahmen der Performanz erfassen. Die standardisierten Faktorladungen der einzelnen Indikatoren stützen diesen Befund. Sie laden alle signifikant auf die latente Variable. 13 der 15 Faktorladungen sind $> .70$ (Backhaus et al., 2018; Weiber & Mülhhaus, 2014). Die Faktorladungen der Indikatoren «Lebensweltlicher Bezug» und «Zeitplanung» weisen allerdings Werte $< .50$ auf. Bei einer standardisierten Faktorladung von .44 des Indikators «Lebensweltlicher Bezug» wird damit nur 20% der Varianz innerhalb des Indikators erklärt. Die Varianz innerhalb des Indikators «Zeitplanung» liegt mit einer standardisierten Faktorladung von .32 bei 10%. Aufgrund der inhaltlichen Bedeutung dieser beiden Indikatoren für die Planung von Unterricht, insbesondere auch von naturwissenschaftlichem Unterricht, werden sie aber vorerst im Messmodell belassen. Die Erkenntnisse, dass mit einer guten Zeitplanung die aktive Lernzeit (Carnahan, 1980), und mit einem sinnvollen Lebensweltbezug das situationale Interesse der Schülerinnen und Schüler, erhöht werden kann (Habig et al., 2018), stützen diesen Entscheid. Das im Projekt KUBeX entwickelte Ratingmanual ist damit faktoriell valide und erfüllt auf der Grundlage der Teilstichprobe ($n = 70$) die meisten der Gütekriterien. Vor diesem Hintergrund bestätigt sich die Forschungsfrage 1, dass sich auf der Grundlage der Daten das fachdidaktische Wissen zur Planung einer Lerngelegenheit bzw. Unterrichtsstunde zum experimentellen Handeln mit Hilfe des hoch inferenten Instruments in Unterrichtsplanungen auf der Grundlage der Daten valide erfassen lässt.

Die erhobenen Daten wurden auch für eine mögliche Validierung eines 2-Faktorenmodells verwendet. Die Ergebnisse werden im nächsten Abschnitt unter Einbezug der Literatur diskutiert.

5.1.2 Faktorielle Validierung des Manuals (Forschungsanliegen 1b)

Die statistischen Analysen zum 2-Faktorenmodell haben gezeigt, dass die Goodness-of-Fit-Indizes alle, der Chi-Quadrat-Test eingeschlossen, im Bereich der in der Literatur empfohlenen Richtwerte (Geiser, 2011) liegen. Sie sind im Vergleich zu den Werten des eindimensionalen Modells besser, was mit Hilfe eines Modellvergleichs statistisch auch

bestätigt werden konnte. Demnach lassen sich zwei Dimensionen aus den Daten extrahieren. Dieser Befund wird auch durch die standardisierten Faktorladungen gestützt. Diese liegen mit Ausnahme von drei Werten alle über .70. Die drei tieferen Korrelationswerte sind im Gegensatz zum 1-Faktormodell ausserdem $> .50$, weshalb sie als hinreichend eingestuft werden können (Backhaus et al., 2018). Die Ergebnisse zeigen, dass insgesamt 13 von 15 untersuchten Indikatoren signifikant nur auf einen der beiden Faktoren laden. Das bedeutet, dass damit eine klare Zuordnung der Indikatoren zu den Faktoren möglich ist. Die beiden Indikatoren «Auswahl» der Kategorie «Ziele» und «Sequenzierung» der Kategorie «Didaktische Strukturierung» laden hingegen als einzige Indikatoren signifikant auf beide Faktoren. Für die Zuordnung ist die höhere Faktorladung in Anlehnung an die Richtwerte (Backhaus et al., 2018; Weiber & Mülhhaus, 2014) vorgezogen worden. Die Zuteilung lässt sich mit Blick auf den Planungsprozess vermutlich so begründen, dass die Auswahl der Ziele primär in der ersten Phase, d.h. in der Analysephase der Planung, von Bedeutung ist, da sie die Ausrichtung der Planung prägt und die weiteren Planungsschritte darauf aufbauen. Im Vergleich dazu ist die Sequenzierung, d.h. das Schaffen einer kohärenten und konsistenten Reihenfolge der Lerninhalte und Ziele in der Experimentiereinheit eher ein Schritt der zweiten Phase, d.h. der Konstruktionsphase der Planung. Hier sollen die Vermittlungsabsichten mit den Kompetenzansprüchen im Fach und den Kompetenzen der Lernenden abgestimmt und die einzelnen Lernschritte in eine sinnvolle Abfolge gebracht werden. Der Indikator «Zeit», mit dem die zeitliche Angemessenheit der in den Planungen beschriebenen Experimentiereinheit eingeschätzt wurde (cf. Kap. I, 3.2.4), hat demgegenüber keine signifikanten Ladungen gezeigt. Er konnte deshalb im Messmodell nicht mehr berücksichtigt werden. Eine mögliche Erklärung wäre, dass es sich hierbei eher um ein Merkmal der Oberflächenstruktur des Unterrichts handelt, das sich mit Hilfe des Ratingmanuals nicht genügend gut erfassen lässt.

Die sich aufgrund der signifikanten Faktorladungen der Indikatoren abzeichnenden beiden Faktoren sind in Abbildung 24 mit Bezug zum Modell der «Didaktischen Rekonstruktion» dargestellt. Es fällt auf, dass fast zwei Drittel aller Indikatoren mit der Dimension «Analyse» korrelieren, während 5 Indikatoren einen Zusammenhang mit der Dimension «Konstruktion» zeigen. Diese Zuteilung der Indikatoren und damit der fachdidaktischen Wissensfacetten lässt sich bei genauerer Betrachtung auch inhaltlich mit den beiden Dimensionen des Planungsprozesses gut vereinbaren (Abb. 24). Die Dimension «Analyse» wird durch 9 Indikatoren aus den Kategorien «Schülervorstellungen», «Fachliche Klärung», «Unterrichtsziele», «Didaktische Strukturierung» und «Reflexivität» modelliert. Es sind dies Indikatoren, die bezogen auf den Planungsprozess wichtige, dem Planungsprozess vorgeschaltete Entscheidungen beinhalten.

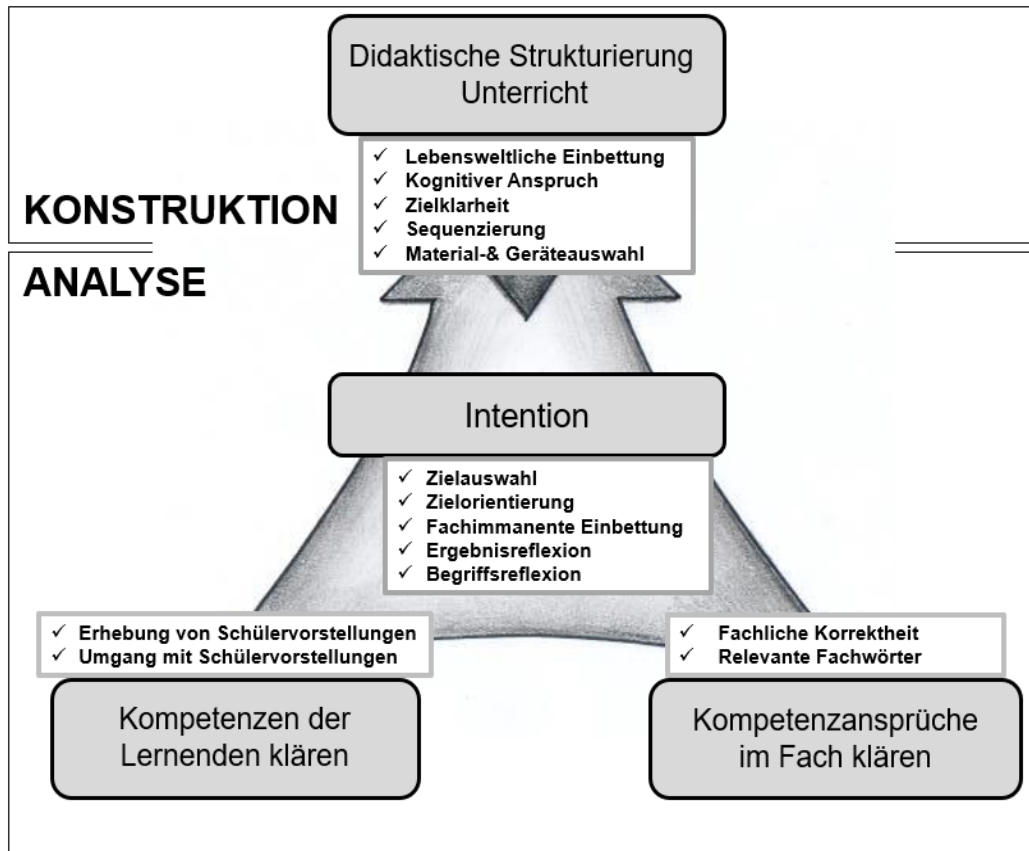


Abbildung 24: 2-Faktorenmodell (Messmodell) mit Bezug zum Planungsmodell der «Didaktischen Rekonstruktion»

Gemäss der Ergebnisse des ESEM laden sowohl die beiden Indikatoren zu den Schülervorstellungen zum experimentellen Handeln («Erhebung» und «Umgang») signifikant hoch und ausschliesslich auf den Faktor «Analyse», als auch zwei der drei Indikatoren der Kategorie «Fachliche Klärung» («Korrektheit», «Relevante Fachwörter»), mit denen die fachliche Richtigkeit des Konzeptes des Teilprozessansatzes und - damit verbunden - die Klärung relevanter Fachwörter in den Daten erhoben worden sind. Auch korrelieren zwei der drei Indikatoren der Kategorie «Ziele» («Auswahl», «Zielorientierung») sowie beide Indikatoren der Kategorie «Reflexivität» («Ergebnisreflexion», «Begriffsbildung») signifikant mit diesem Faktor. Es fällt auch auf, dass der Indikator «Fachimmanente Einbettung», der die inhaltliche Vorbereitung der Schülerinnen und Schüler auf das experimentelle Handeln als Fachinhalt und damit verbunden eine entsprechend bewusste Einbettung des experimentellen Handelns in die Planung erfasst, ebenfalls auf den Faktor «Analyse» lädt. Die hier beschriebenen Indikatoren, die den Faktor «Analyse» abbilden, betreffen demzufolge Wissensfacetten aus den drei zentralen Bereichen «Kenntnisse der Lernenden», «Kompetenzansprüche im Fach» sowie «Vermittlungsabsichten», die im Modell der «Didaktischen Rekonstruktion» enthalten sind und das entsprechende Wissen dazu (van Dijk & Kattmann,

2007; Duit et al., 2012). Besonders erwähnt sei hier noch die Zuordnung der Indikatoren der Kategorie «Reflexivität» zum Faktor «Analyse». Sie deutet darauf hin, dass im Sinne der Sichtbarmachung der Lernprozesse für die Schülerinnen und Schüler (vgl. Börlin, 2012) die Reflexivität im Planungsprozess bewusst analysiert und geplant werden sollte.

Die Dimension «Konstruktion» wird demgegenüber durch 5 Indikatoren aus drei Kategorien abgebildet. Im Zentrum steht die didaktische Strukturierung konkreter und möglichst bedeutsamer Lerngelegenheiten, in denen lernförderliche Korrespondenzen und voraussehbare Lernschwierigkeiten sichtbar werden (Kattmann et al., 1997). Gemäss Lipowsky (2015) scheint eine didaktische Strukturierung mit einer sorgfältigen Planung einherzugehen und eine Voraussetzung für die Anforderungen der Lernenden zu sein. Die 5 Indikatoren, die auf diesen Faktor laden, beschreiben Wissensfacetten, die für didaktische Strukturierung einer qualitätsvollen Lerngelegenheit oder eines Unterrichts auf der Lernprozess-ebene von Bedeutung sind. Dies trifft insbesondere für die Indikatoren «Lebensweltbezug», «Kognitiver Anspruch» und «Sequenzierung» der Kategorie «Didaktische Strukturierung» zu (Baumert et al., 2011; cf. Habig et al., 2018, Hardy et al., 2006). Doch auch die beiden Indikatoren «Zielklarheit» der Kategorie «Ziele» und «Material- und Geräteauswahl» der Kategorie «Fachliche Klärung» stellen inhaltlich strukturierende Massnahmen der Lernunterstützung dar (Adamina et al., 2017). Dabei erfasst der Indikator «Zielklarheit», inwiefern es der Lehrperson in der Unterrichtsplanung gelingt, die Problemstellung und damit das experimentelle Handeln als Fachinhalt den Schülerinnen und Schülern einleuchtend darzustellen, während der Indikator «Material- und Geräteauswahl» die Wahl der Materialien hinsichtlich ihrer Adäquatheit und der intendierten Ziele prüft.

Die aus der faktoriellen Validierung erfolgte Zuordnung lässt sich somit, vor dem Hintergrund der Modelle der «Didaktischen Rekonstruktion» und des ERTE-Modells sowie mit Hilfe der, in Tabelle 5 erfolgten Literaturbezüge, theoretisch und z.T. auch empirisch begründen (Duit et al., 2012; van Dijk & Kattmann, 2007). Das statistisch geprüfte 2-Faktorenmodell ist inhaltlich und faktoriell valide und passt signifikant besser auf die Daten. Das bedeutet, dass das «Fachdidaktische Wissen angehender Lehrpersonen zur Planung von Unterricht zum experimentellen Handeln» im Sinne der Performanz als *PCK on action* (Smith & Banilower, 2015) mit dem Ratinginstrument erfasst werden kann. Auch hat das in Mplus erfolgte ESEM ergeben, dass es innerhalb des Planungsprozesses offenbar zwei Phasen mit Tätigkeiten gibt, die stärker miteinander korrelieren. Die extrahierbaren Phasen sind aufgrund der Zuordnung der Indikatoren gut mit den beiden Dimensionen des Modells der «Didaktischen Rekonstruktion» vereinbar. Hier sind auch Parallelen zum zweidimensionalen Konstrukt von Gess-Newsome et al. (2017), das für die Analyse von Unterricht verwendet worden ist, erkennbar. Allerdings werden im vorliegenden 2-Faktorenmodell

die Kategorien «Schülervorstellungen» und «Reflexivität» eher mit dem Fachwissen und weniger mit dem pädagogischen Wissen in Verbindung gebracht. Das lässt sich damit begründen, dass die Kenntnisse von und der Umgang mit Schülervorstellungen, aber auch die Reflexion zu den Ergebnissen und den Fachbegriffen enger mit dem Fachwissen verknüpft sind. Das Forschungsanliegen 1b, ob sich in den erfassten Daten die zentralen Wissensdimensionen «Analyse» und «Konstruktion» des ERTE-Modells abbilden, lässt sich auf der Datengrundlage der Teilstudie I faktoriell wie auch inhaltlich aufzeigen. Vor diesem Hintergrund kann mit der hier beschriebenen, explorativen Vorgehensweise folgende Hypothese abgeleitet werden: Wenn Unterricht im Sinne der Tiefenstruktur auf Lernprozess-ebene geplant wird, ist der Planungsprozess durch zwei Phasen gekennzeichnet ist, deren Tätigkeiten jeweils eng in Zusammenhang zueinander stehen. Die Ergebnisse sprechen dafür, dass in der Phase der Analyse, bei der es um die Klärung der Kompetenzen der Lernenden, die Klärung der Kompetenzansprüche im Fach und die Vermittlungsabsichten bzw. Intentionen geht, wichtige Grundlagen für die Konstruktion qualitätsvoller Lerngelegenheiten geschaffen werden. Die Konstruktion von Unterricht kann somit nur unter Rückgriff auf die Analyse rekursiv erfolgen. Entsprechend kommt der Analysephase eine wichtige Bedeutung zu.

Wie in Kap. I, 2.4.2 dargelegt, variieren Erhebungsinstrumente in Abhängigkeit zur Erhebungssituation, zum Kompetenzverständnis und zum Ziel der Erhebung (Kaufhold, 2006). Die hier beschriebenen statistischen und inhaltlichen Befunde zum Ratingmanual sollen vor diesem Hintergrund im Hinblick auf eine adäquate Einordnung im nächsten Abschnitt diskutiert werden.

5.2 Befunde zum Design und zur methodischen Herangehensweise in Teilstudie I

Die vorliegende Studie war, wie in Kapitel I, 3.1 erläutert, Teil des binationalen und interdisziplinären Projekts KUBeX, an dem sich vier pädagogische Hochschulen beteiligt haben. Demgemäss ist die Teilstudie I an diesen Pädagogischen Hochschulen im Rahmen der laufenden Ausbildung angehender Lehrpersonen der Sekundarstufe I erfolgt. Zur Erfassung des «Fachdidaktischen Wissens zur Planung von Unterricht zum experimentellen Handeln» wurde eine für Studierende praxisnahe Planungsaufgabe gewählt. Für die Analyse und Einschätzung dieses Wissens konnten neben erstellten Unterrichtsplanungen und -materialien auch videografierte Planungsgespräche genutzt werden. Alle Daten sind nach dem Hypercodingverfahren (Irion, 2010) eventbasiert durch zwei unabhängige Beurteiler eingeschätzt worden. Zur Prüfung der Objektivität des Kodiervorgangs wurden 23

zufällig ausgewählte Unterrichtsplanungen,-materialien und videografierte Planungsgespräche doppelt geratet.

Aufgrund des binationalen Settings des Projekts und der damit verbundenen Beteiligung der vier Pädagogischen Hochschulen an der Untersuchung, aber auch aufgrund der Einbettung der Untersuchung in die laufende Ausbildung, weist diese Studie eine direkte Anbindung an die Lehrpersonenbildung auf. Die Auswahl der Stichprobe ist aufgrund der jeweiligen Praxismodule erfolgt. Alle Studierenden dieser Module wurden einbezogen, eine spezifische Vorauswahl wurde nicht getroffen. Mit der praxisnahen Aufgabenstellung, die die Studierenden bewältigen mussten, liegt der Fokus der Kompetenzerfassung auf der Performanz. Das Design und die methodische Herangehensweise der Teilstudie I zeichnet sich im Vergleich zu Designs, in denen das fachdidaktische Wissen angehender Lehrpersonen mittels Paper-Pencil-Tests erhoben worden ist (e.g. Grossschedl, Neubrand, et al., 2015; Kirschner et al., 2017) durch seine vergleichsweise hohe ökologische Validität aus. Die Ergebnisse deuten weiter darauf hin, dass sich die für die Analyse des fachdidaktischen Wissens der Studierenden erfolgte Nutzung mehrerer Datenquellen als hilfreich erweist. Die in den videografierten Planungsgesprächen erfolgten Präsentationen der Planungen durch die Studierenden ermöglichten weitere Beobachtungen zu subjektiv geprägten, aber unterrichtsrelevanten Lernereignissen in Form von Äusserungen oder Gesten (cf. Irion, 2010; Kreis & Staub, 2012). Dies führte zu einer differenzierteren Einschätzung der Qualität und damit zu einer akurateren Erfassung des fachdidaktischen Wissens der Studierenden (cf. Nijveldt, 2007; Park & Suh, 2015; Stronge, 2006). Erhebungssituation und -setting haben somit Einblick in möglicherweise handlungsrelevante Kompetenzen oder „*competences*“ der Studierenden zur Planung von Unterricht zum experimentellen Handeln gewährt (Blömeke et al., 2015b) und in Anlehnung an Brovelli et al. (2013) damit grundlegende Erkenntnisse zur fachdidaktischen Vorgehensweise und Planung eines solchen Unterrichts ermöglicht.

Die hier diskutierten Ergebnisse basieren allerdings auf einer Untersuchung, die an einer Stichprobe von 70 Studierenden durchgeführt worden ist. Ihre Aussagekraft gilt es vor dem Hintergrund der Stichprobenrepräsentativität zu hinterfragen und die Grenzen dieser Teilstudie I müssen aufgezeigt werden. Der nächste Abschnitt widmet sich diesem Desiderat.

5.3 Grenzen und Aussagekraft der Teilstudie I

Die Kompetenzerfassung in Teilstudie I ist durch ein vergleichsweise aufwändiges Beobachtungsverfahren erfolgt. Die Stichprobengrösse ist mit $N = 119$ aus testökonomischen Gründen, aber auch bedingt durch die jeweiligen Ausbildungssituationen an den vier

Pädagogischen Hochschulen relativ klein. Letzteres hat dazu geführt, dass sich die Stichproben der verschiedenen Pädagogischen Hochschulen in ihrer Grösse z.T. stark unterscheiden (cf. Kap. I, 3.2.2). Aufgrund dieser Gegebenheiten und aus organisatorischen Gründen wurden die wenigen Studierenden der Pädagogischen Hochschule Thurgau der Interventionsgruppe zugeteilt. Die übrigen Studierenden wurden randomisiert der Kontroll- bzw. Interventionsgruppe zugeordnet. Die daraus resultierenden unterschiedlichen Gruppengrössen konnten deshalb nicht vermieden werden. Zur Gewährleistung der Vergleichbarkeit mussten die Daten ausserdem auf der Basis einer Vignette erhoben werden, dies, obgleich die Studie direkt in eine Lehrveranstaltung mit Praxisbezug eingebettet war. Die Studierenden waren auf diese Weise angehalten, den Planungsauftrag losgelöst von anderen Planungsaufträgen umzusetzen und eine Doppellektion zum «Experimentierenlernen am Beispiel der «Visuellen Wahrnehmung» zu gestalten. Ein Teil der Authentizität der Planungssituation ging so verloren. Eine Konfundierung mit anderen Planungsaufträgen, die in den Modulen stattgefunden haben, konnte aufgrund dieses Settings nicht vermieden werden. Es kann daher nicht ausgeschlossen werden, dass es nicht auch Interaktionen mit anderen Planungsaufgaben gegeben hat. Daher sollten die Ergebnisse zur Umsetzung experimentellen Handelns in Unterrichtsplanungen in Teilstudie II vor diesem Hintergrund eingeordnet werden.

Bei der Analyse der Unterrichtsplanungen, -materialien und videografierten Planungsgespräche hat sich weiter herausgestellt, dass das «Fachdidaktische Wissen zur Planung von Unterricht zum experimentellen Handeln» nur in 70 Fällen zumindest ansatzweise beobachtet werden konnte. Der Grund ist darin zu suchen, dass das Instrument zwar das «Fachdidaktische Wissen zur Planung von Unterricht zum experimentellen Handeln» valide erfassen kann (Kap. I, 5.1.1), aber zu wenig eindeutig zwischen dem Vorhandensein in verschiedenen Qualitätsstufen und dem Fehlen dieser Planungskompetenz unterscheidet. Die Konsequenz davon ist ein erheblicher Bodeneffekt, der sich in den Daten zeigt. Um einen solchen Effekt zu vermeiden, hätte dieser Aspekt im Manual berücksichtigt und hätten dafür entsprechende Codes definiert werden müssen. Auch hat sich gezeigt, dass nicht alle Indikatoren in den Unterrichtsplanungen, -materialien und videografierten Planungsgesprächen beobachtet werden konnten. Die in der Vignette aufgeführten Vorgaben haben vermutlich dazu beigetragen, dass sowohl Hinweise zum Lehrplan als auch zu möglichen Individualisierungsmassnahmen in den Daten gefehlt haben.

Die Validierung des Ratinginstruments ist demnach nur auf der Grundlage von 70 Unterrichtsplanungen, -materialien und videografierten Planungsgesprächen durchgeführt worden. Gemäss Bühner (2011) ist eine Stichprobengrösse von $n = 60$ mit Kommunalitätswerten $> .60$ gerade noch ausreichend. Wie in Kap. I, 4.1.3 dargelegt, waren die Kommuna-

litätswerte der Indikatoren in der vorliegenden Untersuchung mehrheitlich gut. Eine erneute Validierung des Instruments an einer anderen, grösseren Stichprobe wäre hier allerdings angezeigt. Auch würde es sich lohnen, die Messgenauigkeit des Ratingmanuals mit einer Generalisierbarkeitsstudie zu prüfen (Cronbach, Rajaratnman & Gleser, 1963), da mit diesem Verfahren mehrere Fehlerquellen (z.B. Kodierer, Unterrichtsplanungen, Indikatoren) nicht voneinander getrennt, sondern gleichzeitig untersucht werden könnten (cf. Park & Suh, 2015).

Zusammenfassend kann festgehalten werden, dass auf der Basis der vorliegenden, gerade ausreichenden Datengrundlage das Ratingmanual das fachdidaktische Wissen zur Planung von Unterricht zum experimentellen Handeln valide erfasst. Ob und in welcher Qualität das experimentelle Handeln tatsächlich in den Unterrichtsplanungen, -materialien und videografierten Planungsgesprächen beobachtet werden konnte, ist Gegenstand der Teilstudie II. Diese soll in der Folge vorgestellt werden.

Teilstudie II

1 Theoretischer und empirischer Hintergrund

Die Handlungskompetenz von Lehrpersonen wird als zentrale Handlungsressource oder *knowledge base of teaching* angesehen (Harms & Riese, 2018; Verloop, van Driel & Meijer, 2001). Sie gilt als wichtige Voraussetzung für erfolgreichen Unterricht (Alonzo & Kim, 2016; Baumert & Kunter, 2006, 2011; Besser & Krauss, 2009; Kirschner et al., 2016; Küsting et al., 2009; Pauli & Reusser, 2009; Wagner, 2014; Olga Zlatkin-Troitschanskaia, Beck, Sembill, Nickolaus & Mulder, 2009b) und wesentlichen Einflussfaktor für die Kompetenzentwicklung von Schülerinnen und Schülern (u.a. Kunter et al., 2013; Lipowsky, 2009). Dabei stehen vor allem kognitive und affektive Dispositionen von Lehrpersonen als Voraussetzung für kompetentes Handeln im Unterricht im Fokus (Baumert & Kunter, 2011b; Ohle et al., 2011; Riese, 2009). In der Literatur besteht Einigkeit darüber, dass im Sinne des Expertenparadigmas das Wissen und Können von Lehrpersonen die zentralen Komponenten ihrer Professionalität darstellen, neben den Einstellungen, Überzeugungen, motivationalen Orientierungen und selbstregulativen Fähigkeiten (Harms & Riese, 2018; Jüttner & Neuhaus, 2013). Für die Ausbildung einer Handlungskompetenz reicht es jedoch nicht aus, wenn eine Lehrperson hohes Professionswissen im Studium erwirbt. Sie muss auch in der Lage sein, dieses Wissen während des komplexen Planungs- bzw. Unterrichtsgeschehens anzuwenden. Es ist aber nach wie vor unklar, welche Dimensionen des Professionswissens Einfluss auf einen guten und lernwirksamen Unterricht haben. Gemäss von Aufschnaiter & Blömeke (2010, S. 365) erfordert unterrichtliches Handeln „ein komplexes Gefüge aus Wissensbeständen (z.B. Schülervorstellungen, Unterrichtskonzeptionen oder der Funktion von Modellen in der Wissenschaft), Fähigkeiten und Fertigkeiten im Bereich der kriteriengeleiteten Unterrichtsplanung, Unterrichtsdurchführung und Reflexion“.

Die nächsten Abschnitte befassen sich vor diesem Hintergrund mit dem Professionswissen als Handlungsressource für (angehende) Lehrpersonen. Dabei werden in einem ersten Schritt empirische Befunde zu Modellierungen des Professionswissens vorgestellt (Kap. II, 1.1). Anschliessend werden zwei Wissensdimensionen, die im Projekt KUBeX im Zentrum gestanden haben (cf. Kap. I, 3.1), nämlich das Fachwissen (Kap. II, 1.1.1) und das fachdidaktische Wissen (Kap. II, 1.1.2), hinsichtlich ihres Einflusses auf die Unterrichtsqualität und die Schülerleistungen näher beleuchtet. Konzeptualisierungen und Operationalisierungen des Fachwissens, welche im Gegensatz zum fachdidaktischen Wissen in Teilstudie I nicht dargelegt worden sind, werden hier noch ergänzt. Kapitel II, 1.2 widmet sich den empirischen Befunden zu den professionellen Kompetenzen der (angehenden) Lehrpersonen zum experimentellen Handeln als Fachinhalt. Diese werden vorgestellt und

eingeordnet. In der Folge befasst sich Kapitel II, 1.3 mit dem Planungshandeln von (angehenden) Lehrpersonen im Kontext der Lehrpersonenbildung. Im Sinne eines Fazits werden dann zentrale theoretische und empirische Erkenntnisse in Kapitel II, 1.4 zusammengefasst. Ausgehend davon gibt Kapitel II, 1.5 einen kurzen Überblick über die Teilstudie II. Das Theoriekapitel schliesst danach ab mit den konkreten Fragestellungen sowie den jeweils entsprechenden, aus den empirischen Befunden abgeleiteten Hypothesen (Kap. II, 1.6).

1.1 Professionswissen – Handlungsressource von (angehenden) Lehrpersonen

Zur Beschreibung der Struktur des Professionswissens hat sich in aktuellen Studien, wie in Kap. I, 2.5.1 beschrieben, meist die von Shulman (1986) und Bromme (1997) begründete Dreiteilung des Wissens etabliert (Baumert & Kunter, 2011a, S. 33; Blömeke, Paine, et al., 2008), die zwischen Fachwissen (CK, *content knowledge*), fachdidaktischem Wissen (PCK, *pedagogical content knowledge*) und pädagogischem Wissen (PK, *pedagogical knowledge*) unterscheidet (cf. von Aufschnaiter & Blömeke, 2010; Borowski et al., 2010; Schmelzing, 2010; Tepner et al., 2012). Während das pädagogische Wissen eher interdisziplinär ausgelegt wird (Voss & Kunter, 2013), werden das Fachwissen und das fachdidaktische Wissen domänenspezifisch aufgefasst (Förtsch, Werner, Kotzebue & Neuhaus, 2016). Wie einige aktuelle Forschungsergebnisse in Naturwissenschaften und Mathematik zeigen, handelt es sich bei diesen drei Wissensdimensionen um drei distinkte Wissensformen, die aber gleichwohl miteinander korrelieren, wenn auch unterschiedlich (cf. Baumert & Kunter, 2011a; Gess-Newsome et al., 2017; Grossschedl, Harms, Kleickmann & Glowinski, 2015; Jüttner & Neuhaus, 2013; Kirschner et al., 2016; Krauss, Brunner et al., 2008; Park & Chen, 2012; Riese & Reinhold, 2012b). Die Stärke der Korrelation scheint demnach test- und stichprobenabhängig zu sein (Smit, Weitzel, et al., 2017). Eine Generalisierung dieser Erkenntnis für die gesamte Population angehender Lehrpersonen ist aufgrund der unterschiedlichen Operationalisierungen dieser Wissensdimensionen, wie auch der unterschiedlichen Fachinhalte kaum möglich (Grossschedl, Harms, et al., 2015). Jedoch ist ein Konsens dahingehend festzustellen, dass diese drei Wissensdimensionen als wesentliche Faktoren professioneller Kompetenz angesehen werden können. Entsprechend sind sie Gegenstand vieler empirischer Forschungsprojekte, wobei gewisse Projekte nur eine Auswahl ins Zentrum ihrer Untersuchung stellen (cf. Brovelli, Rehm & Wilhelm, 2013; Kulgemeyer et al., 2012; Ohle et al., 2011; Schmelzing et al., 2013), andere wiederum alle drei Wissensdimensionen (cf. Baumert & Kunter, 2006; Grossschedl et al., 2015; Kirschner et al., 2016; Tepner et al., 2012). In Bezug auf die drei erwähnten Dimensionen weisen Stiller et al. (2014) ausserdem noch darauf hin, dass der aus dem Englischen übersetzte Begriff

Wissen für *knowledge* zu kurz greift. Eine Unterscheidung in eher deklarative Bestandteile eines Professionswissens und Bestandteile fachwissenschaftlicher, fachdidaktischer und pädagogischer Kompetenzen (die Fähigkeit, dieses Wissen anzuwenden) sei angemessener.

Aufgrund der unterschiedlichen Auffassungen und Ansätze bei den Konzeptualisierungen und Operationalisierungen des Professionswissens bleibt weiterhin unklar, welche Bereiche des Professionswissens in welcher Ausprägung notwendig sind, damit eine Lehrperson erfolgreich unterrichten kann (von Aufschnaiter & Blömeke, 2010). Bisherige Arbeiten sind vornehmlich in der Mathematik durchgeführt worden. In der naturwissenschaftsdidaktischen Forschung ist das Professionswissen erst in den letzten Jahren vermehrt in den Fokus gerückt (Stiller et al., 2014). Die Erkenntnislage ist daher vergleichsweise geringer und theoriegeleitete Untersuchungen des breiten Kompetenzspektrums entsprechend selten (Riese & Reinhold, 2012b). Auch zur Frage, inwiefern das fachdidaktische Wissen und das Fachwissen die Unterrichtsqualität und auch die Schülerleistungen beeinflussen, ist die Datenlage nach wie vor inkonsistent (Abell, 2007). Die nächsten beiden Abschnitte widmen sich dieser Sachlage. Da im Theorieteil der Teilstudie I die Konzeptualisierungen und Operationalisierungen des Fachwissens nicht dargelegt worden sind, werden diese hier noch ergänzt.

1.1.1 Das Fachwissen (*content knowledge*, CK)

Es ist nicht von der Hand zu weisen, dass Fachunterricht ohne Fachwissen nicht möglich ist. „Allerdings ist nicht ganz klar, welches Fachwissen tatsächlich dafür benötigt wird“ (Borowski, Kirschner, Liedtke & Fischer, 2011, S. 4). Entsprechend sind in der Literatur unterschiedliche Konzeptualisierungs- und Klassifizierungsansätze zum notwendigen Fachwissen (*content knowledge*, CK) zu finden, über das Lehrpersonen verfügen sollten (Baumert et al., 2010, 2011; Blömeke, Kaiser, et al., 2008; Blömeke, Kaiser & Lehmann, 2010; Riese & Reinhold, 2010). Der Handlungsrahmen der Lehrtätigkeit jedoch wird grundsätzlich durch das Schulfach vorgezeichnet. Demgemäss ist in der Literatur ein Konsens erkennbar, wonach es zur adäquaten fachlichen Vorbereitung des Unterrichts ein Schulfachwissen sowie ein vertieftes, konzeptionelles Schulfachwissen braucht, das über die Schulstufen hinausgeht (Ball, Thames & Phelps, 2008; Baumert et al., 2010; Blömeke et al., 2009; Blömeke & Müller, 2009; Kirschner et al., 2016; Krauss et al., 2011; Tepner et al., 2012) und domänenspezifisch ist (Neuhaus, 2007; Seidel & Shavelson, 2007; Tepner et al., 2012).

Es gibt Hinweise, wonach das Niveau des Fachwissens einen mediiierenden Einfluss darauf hat, wie der Fachinhalt unterrichtet wird (Carlsen, 1991; Gess-Newsome & Lederman, 1995). Der Lehrperson kommt eine besondere Verantwortung in der Auswahl und Schwerpunktsetzung der Fachinhalte zu. Shulman beschreibt diese so:

„The teacher has special responsibilities in relation to content knowledge, serving as the primary source of student understanding of subject matter. The manner in which that understanding is communicated conveys to students what is essential about a subject and what is peripheral” (Shulman, 1987, S. 9).

Für den Biologieunterricht gibt es in Bezug auf die Auswahl der Inhaltsbereiche im internationalen Raum noch keine Einigkeit (Jüttner & Neuhaus, 2013). Dies widerspiegelt sich in aktuellen Untersuchungen zum Fachwissen (CK) in den Naturwissenschaften. Zum einen sind diese domänenübergreifend angelegt, wie das Projekt ProwiN (Borowski et al., 2010; Tepner et al., 2012) oder das KiL-Projekt (Grossschedl & Harms, 2013; Kröger, Euler, Neumann, Härtig & Petersen, 2012), zum anderen domänenspezifisch (e.g. Kulgemeyer, 2013; Riese & Reinhold, 2010). Das Fachwissen ist in diesen Projekten allerdings immer domänenspezifisch modelliert. Stiller et al. (2014) unterscheiden innerhalb des Fachwissens mit Verweis auf Grossman, Wilson, & Shulman (1989) ausserdem verschiedene Facetten. Sie unterteilen das Fachwissen in ein allgemeines Wissen zum Fachinhalt eines Fachgebietes und in zwei weitere Wissensfacetten. Letztere bezeichnen sie mit *substantive* und *syntactic knowledge* (cf. Grossman, 1990, S. 6f.). Als *substantive knowledge* erachten sie die Struktur und Organisation des konzeptionellen Wissens, die letztlich die Fragestellungen weiterer Untersuchungen beeinflusst. Unter *syntactic knowledge* inkludieren sie u.a. das Wissen über die fachlichen Methoden. Damit verbunden ist auch ein Wissen darüber, was in der *Community* als Evidenz anerkannt wird und nach welchen Regeln Erkenntnisse gewonnen und evaluiert werden (Grossman, 1990, S. 6). Diese Unterscheidung wird auch in der Arbeit von Kotzebue & Nerdel (2012) zum Professionswissen von Biologielehrkräften im Umgang mit Diagrammen vorgenommen und erscheint im Kontext des vorliegenden Forschungsvorhabens besonders erwähnenswert.

In der gleichen Weise, wie sich die Konzeptualisierungen und Operationalisierungen des Fachwissens unterscheiden, erweist sich die empirische Befundlage zu den Auswirkungen des Fachwissens auf den Unterricht und die Schülerleistungen nach wie vor als inkonsistent (Abell, 2007). Ein Grund dafür ist in der unterschiedlichen methodischen Herangehensweise und Evaluation zu suchen (Abell, 2007; Jüttner, Boone, Soonhye & Neuhaus, 2013). So wurde das Fachwissen zum einen anhand distaler Faktoren erfasst, beispielsweise der Qualifikation einer Lehrperson oder der Anzahl Aus- und Weiterbildungskursen, die eine

Lehrperson absolviert hat (Baumert et al., 2010). Zum anderen wurde das Fachwissen proximal, d.h. direkt und anforderungsbezogen u.a. über Tests erhoben (cf. Jüttner et al., 2013). Entsprechend divers sind die Befunde. Während Hill, Rowan & Ball (2005) einen direkten Zusammenhang zwischen dem Fachwissen und den Leistungen der Schülerinnen und Schüler in Mathematik in der Grundschule nachweisen konnten, war ein solcher Effekt in anderen Studien nicht feststellbar (Ohle, 2010; Ohle et al., 2011). In gewissen Studien konnte aber ein indirekter Effekt nachgewiesen werden, nämlich über die curriculare Passung von Aufgaben (Baumert et al., 2010) oder über die Sequenzierung der Lernprozesse, eines Qualitätskriteriums von Unterricht (Ohle et al., 2011). In einer Pilotstudie von Lagler & Wilhelm (2013) erwies sich die Fachlichkeit in der Ausbildung von Lehrpersonen als Prädiktor für die Leistung der Lernenden in den Fächern Chemie und Physik, nicht aber in der Biologie. Die Autoren erklären dies u.a. damit, dass in der Biologie offenbar eine schlechte Passung zwischen der universitären Fachausbildung und den im Lehrplan vorgegebenen Fachinhalten bestand. Daraus abgeleitet schreiben sie dem effektiven, unterrichtsrelevanten Wissen von Lehrpersonen eine grössere Bedeutung zu. Befunde früherer Studien weisen ebenfalls darauf hin, dass das Niveau des Fachwissens einen Einfluss darauf hat, wie der Fachinhalt unterrichtet wird (Carlsen, 1991; Gess-Newsome & Lederman, 1995). Kunter et al. (2011) sehen das in gewisser Weise ähnlich. Sie erachten das Fachwissen als notwendige, aber nicht hinreichende Voraussetzung für qualitätvollen Unterricht und für die Lernfortschritte bei den Schülerinnen und Schülern (cf. Abell, 2007). Aus ihrer Perspektive „definiert das Fachwissen den Entwicklungsraum des fachdidaktischen Wissens und damit indirekt auch die Unterrichtsqualität“ (Baumert & Kunter, 2011, S. 185), wobei das im Unterricht verfügbare fachdidaktische Handlungsrepertoire weitgehend von der Breite und Tiefe des konzeptuellen Fachverständnisses abhängt (Baumert & Kunter, 2011, S. 166). Ergebnisse aus dem Verbundprojekt ProwiN haben hierzu aktuell ergeben, dass das fachdidaktische und das fachliche Wissen bei allen drei naturwissenschaftlichen Fächern mehrheitlich signifikant korrelieren, dass aber die Korrelationen zwischen dem fachdidaktischen und pädagogischen Wissen mit Ausnahmen kleiner sind und dass erwartungsgemäss keine signifikanten Korrelationen zwischen dem fachlichen und dem pädagogischen Wissen zu verzeichnen sind (Kirschner et al., 2017, S. 121). Grossschedl et al. (2015) interpretieren den Zusammenhang zwischen PCK und CK in ähnlicher Weise wie Kunter et al. (2011) und sehen im Fachwissen die Relevanz für die Entwicklung fachdidaktischen Wissens. Letzteres soll im nächsten Abschnitt hinsichtlich seiner Wirksamkeit auf den Unterricht und die Leistungen der Schülerinnen und Schüler beleuchtet werden.

1.1.2 Das fachdidaktische Wissen (*pedagogical content knowledge*, PCK)

Die Vermutung, dass das fachdidaktische Wissen der Lehrperson Einfluss auf die Qualität des Unterrichts und somit die Schülerinnen- und Schülerleistungen nimmt, konnte in Teilen durch aktuelle Befunde bereits gestützt werden. So wies die LMT-Studie (*Learning mathematics for teaching*; Hill et al., 2005) nach, dass deklaratives fachdidaktisches Wissen einer Lehrperson ein Prädiktor für Schülerleistung im Mathematikunterricht darstellt. Auch die COACTIV-Studie gelangte zu ähnlichen Ergebnissen (Baumert et al., 2010). Ausserdem zeigte sie eine mittlere bis hohe Korrelation zwischen dem Fachwissen und dem fachdidaktischen Wissen (Krauss, Neubrand, et al., 2008) und konnte ausweisen, dass das fachdidaktische Wissen die kognitive Struktur der im Unterricht verwendeten Lerngelegenheiten bestimmt. Letzteres wurde auch in Studien zum fachdidaktischen Wissen von Physiklehrpersonen von Olszewski (2010) und Ergönenc, Neumann, & Fischer (2014) festgestellt. Während Olszewski (2010) zudem noch einen korrelativen Zusammenhang zwischen dem fachdidaktischen Wissen der Lehrpersonen und den Schülerergebnissen ermittelte, konnte dieser in der Studie von Ergönenc, Neumann & Fischer (2014) nicht nachgewiesen werden. Eine Untersuchung im Rahmen des naturwissenschaftlichen Sachunterrichts verzeichnete ebenfalls einen positiven Zusammenhang zwischen dem fachdidaktischen Wissen der Lehrpersonen und dem Lernerfolg der Lernenden (Lange, Kleickmann, Tröbst & Möller, 2012). Obgleich hier einige neuere Ergebnisse zum Einfluss des fachdidaktischen Wissens von Lehrpersonen auf Unterricht und Schülerleistungen festzustellen sind, ist der Forschungsstand angesichts der zentralen Bedeutung dieser Wissensdimension für die Gestaltung qualitätvollen Unterrichts nach wie vor unbefriedigend (Depaepe et al., 2013; Schneider & Plasman, 2011). Auch bleibt ungeklärt, ob diese Ergebnisse auf die anderen naturwissenschaftlichen Fächer, insbesondere auf die Biologie übertragen werden können und was dies für die Ausbildung angehender Biologielehrpersonen bedeutet. Gemäss Schmelzing, Wüsten, Sandmann und Neuhaus (2008) finden sich für das Unterrichtsfach Biologie bisher kaum Studien zum Einfluss des Professionswissens, insbesondere des fachdidaktischen Wissens von Biologielehrkräften auf die Unterrichtsqualität und Schülerleistungen im Biologieunterricht. Es sind seither einige Studien dazugekommen, die sich diesem Desiderat in der einen oder anderen Form gewidmet haben. In einer Studie zum Unterricht von Biologielehrpersonen konnten Park, Jang, Chen und Jung (2011) zeigen, dass das fachdidaktische Wissen mit der Qualität des Unterrichts korreliert. Aufgrund dieser Ergebnisse vertreten sie die Ansicht, dass „*teaching excellence should include the evaluation of PCK, since it appears that PCK is a reliable predictor of what a teacher knows and what a teacher is actually doing in the classroom*“ (Park et al., 2011, S. 253). Vergleichs-

weise ähnliche Erkenntnisse wurden in einer Studie von Brühwiler (2014) berichtet. In dieser Studie erreichten Schülerinnen und Schüler bei adaptiven Lehrpersonen einen grösseren Wissenszuwachs, wobei dieser Effekt nicht direkt, sondern moderiert über die Unterrichtsqualität, vermittelt wurde. Gess-Newsome et al. (2017) gelang es im Gegensatz dazu nicht, einen solchen Zusammenhang in ihrer Studie nachzuweisen. Sie konnten lediglich einen statistisch signifikanten Zusammenhang zwischen dem pädagogischen Wissen und der Unterrichtspraxis feststellen. Auch der erhoffte Zusammenhang zwischen einem ausgeprägten fachdidaktischen Wissen und guten Schülerleistungen konnte aufgrund der erhobenen Daten nicht ermittelt werden. Henze und van Driel (2015) führen diese fehlende Korrelation darauf zurück, dass sich das fachdidaktische Wissen je nach Erhebungsart, entweder mit Interviews oder aufgrund der direkten Handlung, stark unterscheidet. In weiteren Studien wie KiL (Grossschedl, Harms et al., 2015), ProwiN (Jüttner et al., 2013), aber auch Arbeiten von Kotzebue und Nerdel (2015) wurden Testinstrumente zur Erfassung fachdidaktischen Wissens konzipiert und mittels Tests (*PCK on action*) geprüft (cf. Park & Suh, 2015). Implikationen des fachdidaktischen Wissens bzw. des Professionswissens insgesamt auf das Unterrichtshandeln (*PCK in action*) und die Schülerleistungen sollen hier weiter analysiert werden (Kirschner et al., 2017).

Diese Ausführungen machen deutlich, dass konsistente Beschreibungen, ob und wie sich das Professionswissen der Lehrperson ganz grundsätzlich über ihr konkretes Unterrichtshandeln auf die Unterrichtsqualität und die Lernergebnisse der Schülerinnen und Schüler auswirkt, nach wie vor fehlen (Borowski et al., 2010; Kirschner et al., 2017). Für eine umfassende Kompetenzmessung wäre dies aber unabdingbar. Nach Oser, Heinzer und Salzmann (2010) lassen sich mit den von Shulman (1987, 1986) vorgeschlagenen Wissensdomänen nämlich keine umfassenden Kompetenzmessungen vollziehen, weil der Fokus ausschliesslich auf dem Wissen liegt. Auch kritisieren sie Diagnoseinstrumente, die Kompetenzen unabhängig von situativen Kontexten erfassen. Hasse, Joachim, Bögeholz und Hammann (2014) deuten in ihrer Arbeit darauf hin, dass Wissen und Kompetenzen zwei verschiedene Konstrukte darstellen. Während Kompetenzen den Fokus auf die Problemlösefähigkeiten und -fertigkeiten legen (Weinert, 2001b), stellt das Wissen eine notwendige und wichtige Grundlage für die Erlangung einer Kompetenz dar (Weinert, 1999). Vogelsang und Reinhold (2013, S. 116) betonen, dass ohne die Betrachtung der Handlungsebene die Handlungsrelevanz von erfasstem Professionswissen nicht eindeutig begründbar ist. Zur Einschätzung des Einflusses auf die Lernenden ist ihrer Meinung nach wirkliches Unterrichtshandeln notwendig. Gemäss Stender, Brückmann und Neumann (2015, 2017) kann vermutet werden, dass explizites Wissen im Unterrichtsplanungsprozess wirksam wird. Hinsichtlich der Lehrerexpertise ist allerdings denkbar, dass das Fachwissen

bzw. das fachdidaktische Wissen zwar eine notwendige, aber nicht hinreichende Disposition für ein erfolgreiches Handeln im Unterricht darstellt (Abell, 2007; Fischler, 2008; Lederman & Gess-Newsome, 1992; Neuweg, 2014; Oser & Oelkers, 2001). Das Können einer Lehrperson sollte jedoch mit der Betrachtung der Performanz, d.h. dem beobachtbaren Verhalten, feststellbar sein und sich in der Qualität des von ihr (geplanten oder) realisierten Unterrichts zeigen (Vogelsang & Reinhold, 2013). Idealerweise müsste daher das Können einer Lehrperson mittels prototypischer Anforderungssituationen, die sie erfolgreich absolvieren muss, erfasst und in Bezug zum erfassten Professionswissen gesetzt werden können (ebd.). Bisher existieren wenige Studien, die Bezüge zwischen dem Professionswissen der Lehrpersonen, ihrem Unterricht und Schülervariablen herstellen und damit beide Forschungsbereiche kombinieren (Cauet, 2016; Förtsch et al., 2016).

Im Projekt KUBeX wurde diese Herangehensweise für die Untersuchung professioneller Handlungskompetenzen angehender Lehrpersonen zum experimentellen Handeln im Sinne der Erkenntnisgewinnung zumindest teilweise gewählt. Dabei wurden das Professionswissen (Fachwissen und fachdidaktisches Wissen) mit Hilfe von Tests erfasst und die Performanz dieses Wissens in Unterrichtsplanungen analysiert. Bisherige empirische Befunde zu den hier untersuchten Kompetenzen von (angehenden) Lehrpersonen zur Planung und Durchführung von Unterricht zum Experimentieren im Sinne der Erkenntnisgewinnung werden im nächsten Abschnitt vorgestellt.

1.2 Professionelle Kompetenzen zum Experimentieren im Sinne der Erkenntnisgewinnung

Die Vermittlung von Kompetenzen zum experimentellen Handeln wird als wichtiges Bildungsziel (EDK, 2011; KMK, 2005, NCR, 2012) angesehen und auch als fachspezifisches Qualitätsmerkmal des Naturwissenschaftsunterrichts eingestuft (cf. Kap. I, 2.6.2). Damit verbunden ist der Aufbau eines Verständnisses der Prozesse, mit denen naturwissenschaftliche Erkenntnisse generiert werden, und derer Grundlage, die durch Konventionen und ethische Grundsätze gekennzeichnet ist (Roberts, 2007; Schwartz, Lederman & Crawford, 2004). Die weit verbreitete Annahme, dass sich ein solches Verständnis durch häufigere Experimentieranlässe im Sinne von „*doing science*“ von selbst aufbaut, hat sich nicht bestätigt (Jerrim, Oliver & Sims, 2019; Schwartz et al., 2004; Windschitl, Thompson & Braaten, 2008). Neben konkreten, authentischen Experimentiersituationen bedarf es unterstützender Massnahmen (Furtak et al., 2012; Jerrim et al., 2019; Lazonder & Harmsen, 2016) sowie expliziter Reflexionsanlässe, die die Lernenden beim Aufbau eines solchen Verständnisses unterstützen (Abd-El-Khalick & Lederman, 2000; Abrahams & Reiss, 2012; Capps & Crawford, 2013a; Fazio et al., 2010; Melville, Fazio, Bartley & Jones, 2008;

Yoon, Joung & Kim, 2012). Bartholomew, Osborne und Ratcliffe (2004) erachten die Art, wie im Unterricht über den Prozess der Erkenntnisgewinnung gesprochen wird, als entscheidenden Faktor für die Gestaltung eines wirksamen Unterrichts. Tesch & Duit (2004) sehen in der Einbettung des Experiments in den gesamten Unterricht einen wichtigen Qualitätsfaktor. Für die Gestaltung eines solch komplexen Unterrichts sollten (angehende) Lehrpersonen über ein fundiertes Professionswissen und die damit verbundenen professionellen Handlungskompetenzen verfügen (Crawford, 2000, 2007), insbesondere auch im Biologieunterricht. In der Vergangenheit hat sich allerdings gezeigt, dass Lerngelegenheiten, die das Erlernen der Teilkompetenzen wissenschaftlichen Denkens im Unterricht fördern könnten, insgesamt gering ausgeprägt sind (Capps & Crawford, 2013a; Fuccia, Witteck, Markic & Eilks, 2012; Hofstein & Lunetta, 2004; Schulz, 2011; Schulz, Riess, Wirtz & Barzel 2012, S. 333). Wenn naturwissenschaftliche Arbeitsweisen wie beispielsweise das Experimentieren im Sinne der Erkenntnisgewinnung im Unterricht eingesetzt werden, dann in der Regel um die Schülerinnen und Schüler für Naturwissenschaften zu motivieren, Phänomene darzustellen, fachliche Inhalte und Konzepte zu erarbeiten oder das praktische Arbeiten zu fördern (Gyllenpalm & Wickman, 2011a; Gyllenpalm, Wickman & Holmgren, 2010; Schneider & Plasman, 2011; cf. Tesch & Duit, 2004; Wallace & Kang, 2004). Die Vermittlung von Teilkompetenzen wissenschaftlichen Denkens, wie z.B. das Formulieren von Fragen oder Hypothesen, das Planen und Durchführen von Experimenten oder auch die Analyse und Interpretation von Daten, ist selten zu beobachten. Die Gestaltung von Experimentalstunden im Sinne der Erkenntnisgewinnung scheint demnach erhebliche Herausforderungen an die Lehrenden zu stellen. Das bestätigen u.a. auch Fazio et al. (2010), die zeigen konnten, dass angehende Lehrpersonen in den Schulpraxisphasen Mühe bekunden, einen Unterricht zu planen, mit dem Kompetenzen naturwissenschaftlicher Erkenntnisgewinnung gefördert werden. Anderson (2002) unterteilt die empirischen Befunde zu den in diesem Zusammenhang bestehenden Herausforderungen in drei Dimensionen, eine technische, eine politische und eine kulturelle Dimension. Den letzten beiden Dimensionen weist er äussere Faktoren zu, die Lehrpersonen bei der Gestaltung eines solchen Unterrichts beeinflussen können (u.a. Qualität der Lehrpersonenbildung, kulturell geprägte Überzeugungen und Werthaltungen etc.). Unter der ersten Dimension subsumiert er Herausforderungen, die hauptsächlich auf den Fähigkeiten und Fertigkeiten der Lehrpersonen gründen und sich im unterrichtlichen Handeln niederschlagen können. Dazu gehören zusammengefasst das fachliche Wissen (*syntactic knowledge*) und Verständnis des Konzepts *Scientific Inquiry*, das fachdidaktische Wissen einen entsprechenden Unterricht zu gestalten, die eigenen Erfahrungen in Bezug auf die Durchführung naturwissenschaftlicher Untersuchungen sowie Überzeugungen zur Lehrtätigkeit, aber auch zum Prozess der Erkenntnisgewinnung (cf. Roehrig & Luft, 2004). Hinsichtlich

des Fachwissens haben empirische Befunde gezeigt, dass (angehende) Lehrpersonen über ein unzureichendes fachmethodisches Verständnis und Wissen (*syntactic knowledge*) zu den Teilkompetenzen einer naturwissenschaftlichen Untersuchung verfügen (Capps & Crawford, 2013b; Chabalengula et al., 2012; Davis et al., 2006; Kunz, 2011). Ähnliche Erkenntnisse wurden auch von Gyllenpalm & Wickman (2011a) und Schneider & Plasman (2011) berichtet, die beobachtet haben, dass Lehrpersonen ihren Unterricht u.a. aufgrund von *Hands-on*-Aktivitäten als „*inquiry-oriented*“ einstufen, obgleich dieses Ziel effektiv nicht verfolgt. Eine weitere Studie von Gyllenpalm et al. (2010) hat ergeben, dass die beiden Konzepte „*Inquiry* als Unterrichtsmethode“ und „*Inquiry* als Methode der Erkenntnisgewinnung“ von angehenden Lehrpersonen oft vermischt wurden. Eine ähnliche Beobachtung ergab sich auch bei der Umsetzung von *Scientific Inquiry*, wie eine Untersuchung zum Umgang mit dem Kernbegriff Hypothese bei praktizierenden Lehrpersonen zeigte (Gyllenpalm & Wickman, 2011a). Crawford (2007) stellte fest, dass angehende Lehrpersonen mit unzureichenden Fachkenntnissen zum Experimentieren im Sinne der Erkenntnisgewinnung Mühe bekunden die Lernenden zu animieren, Fragestellungen oder aber auch Erklärungen zu formulieren (cf. Windschitl, 2003, 2004). In diesem Zusammenhang konnte Gomez-Zwiep (2008) auch zeigen, dass es Lehrpersonen schwerfällt, die Präkonzepte der Schülerinnen und Schülern in ihrem Unterricht zu berücksichtigen und diesen entsprechend auszurichten. Passend dazu merken Wallace und Kang (2004) an, dass der Unterricht massgeblich von den Konzepten abhängt, die Lehrpersonen zum Prozess der Erkenntnisgewinnung vorweisen. Alle hier beschriebenen empirischen Befunde deuten darauf hin, dass (angehende) Lehrpersonen in der Begleitung des Erkenntnisprozesses fachliche und auch fachdidaktische Defizite aufweisen (cf. Lee & Luft, 2008; Yoon et al., 2012), welche die Kompetenzentwicklung bei Schülerinnen und Schülern potenziell behindern können. Allerdings haben Mugaloglu und Saribas (2010) festgestellt, dass die von ihnen untersuchten angehenden Lehrpersonen trotz guter Kenntnisse zum Konzept *Scientific Inquiry* keinen entsprechenden Unterricht geplant haben. Daraus lässt sich ableiten, dass auf der Ebene der Lehrenden auch andere Gründe für die Planung eines solchen Unterrichts von Bedeutung sind. Zum einen können diese, wie Windschitl (2004) darlegt, Capps und Crawford (2013a) aber nicht bestätigen konnten, auch in Zusammenhang mit eigenen Erfahrungen in der Durchführung von naturwissenschaftlichen Untersuchungen stehen. Zum anderen weisen Ergebnisse von Crawford (2007) und Kang (2008) darauf hin, dass auch persönliche Überzeugungen zum Konzept *Scientific Inquiry* und zu dessen Vermittlung die Gestaltung der Unterrichtsplanungen und auch des Unterrichts beeinflussen (cf. Davis et al., 2006).

Hinsichtlich des Lehrens von Kompetenzen wissenschaftlichen Denkens und Arbeitens bestehen bei Studierenden und auch erfahrenen Lehrpersonen demnach Defizite. Wilson, Taylor, Kowalski & Carlson (2010) betonen, dass vor diesem Hintergrund der Professionalisierung in der Ausbildung eine bedeutende Rolle für die Gestaltung eines solchen Unterrichts zukommt. Es bedarf daher geeigneter Anlässe, mit deren Hilfe dieser Professionalisierungsprozess angehender Lehrpersonen unterstützt und das dafür notwendige fachdidaktische Wissen aufgebaut werden kann. Forschungsbefunde zur Planungskompetenz von Lehrpersonen geben deutliche Hinweise, dass das berufliche Handeln durch Unterrichtsplanung verbessert werden kann, da sie handlungsleitend ist und die Handlungssituation strukturiert (König et al., 2015; Weingarten & van Ackeren, 2017). Demzufolge widmet sich der nächste Abschnitt dem Planungshandeln und der damit verbundenen Möglichkeit der Wissensanwendung und Kompetenzentwicklung im Kontext der naturwissenschaftsdidaktischen Ausbildung von Lehrpersonen.

1.3 Das Planungshandeln im Kontext der naturwissenschaftsdidaktischen Lehrpersonenbildung

Das Planungshandeln stellt im Kontext der naturwissenschaftsdidaktischen Ausbildung von Lehrpersonen ein wichtiges Element unterrichtlicher Praxis dar und bietet einen geeigneten Anlass, das fachdidaktische Wissen in einer handlungsnahen Situation anzuwenden (Gess-Newsome, 2015, cf. Kap. I, 2.5.2) und aufzubauen (Brown, Friedrichsen & Abell, 2013). Durch das Planen von Unterricht muss die Lehrperson bewusst themenspezifisches Professionswissen (TSPK) in Handlungsskripte überführen (Stender et al., 2017, cf. Kap. I, 2.8.2). Die Kompetenz des Planungshandelns manifestiert sich somit in der Qualität der entwickelten Handlungspläne. Allgemein wird aber davon ausgegangen, dass die Analyse von schriftlichen Unterrichtsplanungen Aussagen über die Qualität von Unterrichtskonzeptionen als Antizipation einer Handlungssituation ermöglichen (Davies & Rogers, 2000; König et al., 2015; Seel, 1997, 2011; Tillema, 2009). Es muss dabei aber immer berücksichtigt werden, dass der Planungsprozess vielschichtig ist und von den Akteuren meist nicht vollständig verbalisiert wird (Seel, 2011). Ausserdem gewährt die Analyse von Unterrichtsplanungen mit Hilfe von Ratingmanuals nur einen limitierten Einblick in das fachdidaktische Wissen von (angehenden) Lehrpersonen, da es nie möglich ist, alle Aspekte vollständig zu erfassen (Park & Suh, 2015). Das erfassbare Wissen entspricht demzufolge auch nie vollumfänglich dem tatsächlichen Wissen (ebd., S. 115). Zierer, Werner und Wernke (2015) stellen hierzu ausserdem fest, dass Planungsraster im Gegensatz zu Planungsmodellen bei Berufsanfängerinnen und Berufsanfängern zu weniger Planungsüberlegungen führen und damit stärker einschränken. John, (2006) sieht dies in gewisser Weise

ähnlich, da seines Erachtens der kreative, problemlösende Aspekt des Planungsprozesses durch starre Vorlagen verloren geht.

Vor dem Hintergrund, dass Operationalisierungen der Planungskompetenz bisweilen ganz allgemein, aber auch hinsichtlich der Kompetenzstufungen noch ungeklärt sind (Zierer, Werner et al., 2015, S. 378), wurden bisher noch wenige Studien mit einem naturwissenschaftsdidaktischen Fokus durchgeführt, die Aussagen zur Planungskompetenz und damit zur Qualität der Unterrichtsplanungen erlauben (cf. Kap. I, 2.8.3). Eine der wenigen Studien, die sich aus naturwissenschaftsdidaktischer Perspektive mit der Qualität von Unterrichtsplanungen auseinandergesetzt, diese aber über Planungsvignetten erfasst hat, ist eine Studie von Stender et al. (2017). Das Ziel dieser Studie war die Klärung, ob die oben beschriebene Transformation von TSPK (*topic-specific professional knowledge*) zu PPCK (*personal pedagogical content knowledge*) in Unterrichtsplanungen zu einem Physikthema tatsächlich sichtbar wird. Das Erkenntnisinteresse hinsichtlich des themenspezifischen Professionswissens basierte auf Befunden aus einer früheren Studie (Stender, 2014; cf. Stender et al., 2015). Diese konnte bei angehenden Lehrpersonen keinen direkten Einfluss des fachdidaktischen Wissens auf die Qualität der Unterrichtsskripts nachweisen, allerdings aber einen direkten Zusammenhang zwischen der Motivation sowie der selbstregulativen Fähigkeiten und der Qualität der Skripts. Nur bei Referendarinnen und Referendaren konnte ein Einfluss des fachdidaktischen Wissens auf die Qualität der Unterrichtsskripts ermittelt werden, wobei dieser durch affektive Dispositionen moderiert wurde. Gemäss der Autorenschaft schien dieser Befund konform mit der Annahme, dass angehende Lehrpersonen, aufgrund mangelnder Gelegenheiten solche Transformationsprozesse durchzuführen, über ein weniger ausgeprägtes fachdidaktisches Wissen verfügen, aber auch, dass affektive Dispositionen die Qualität von Unterrichtsskripten beeinflussen (Stender et al., 2015). Das Transformationsmodell konnte damals mit diesen Befunden für das Fach Physik bestätigt werden (ebd.). In der darauf aufbauenden aktuellen Studie, an der erfahrene Physiklehrpersonen beteiligt waren, wurde das Transformationsmodell auf der Basis des Konsensmodells (Berry et al., 2015; cf. Kap. 2.5.2) angepasst und ebenfalls geprüft (Stender et al., 2017). Das neue Instrument erwies sich ebenfalls als valide. Zudem konnte ein direkter Einfluss des themenspezifischen Professionswissens auf die Qualität des in den Unterrichtsplanungen beobachtbaren, personenspezifischen, kontextgebundenen PPCK festgestellt werden. Dieser Effekt wurde auch hier durch volitionale und motivationale Dispositionen moderiert.

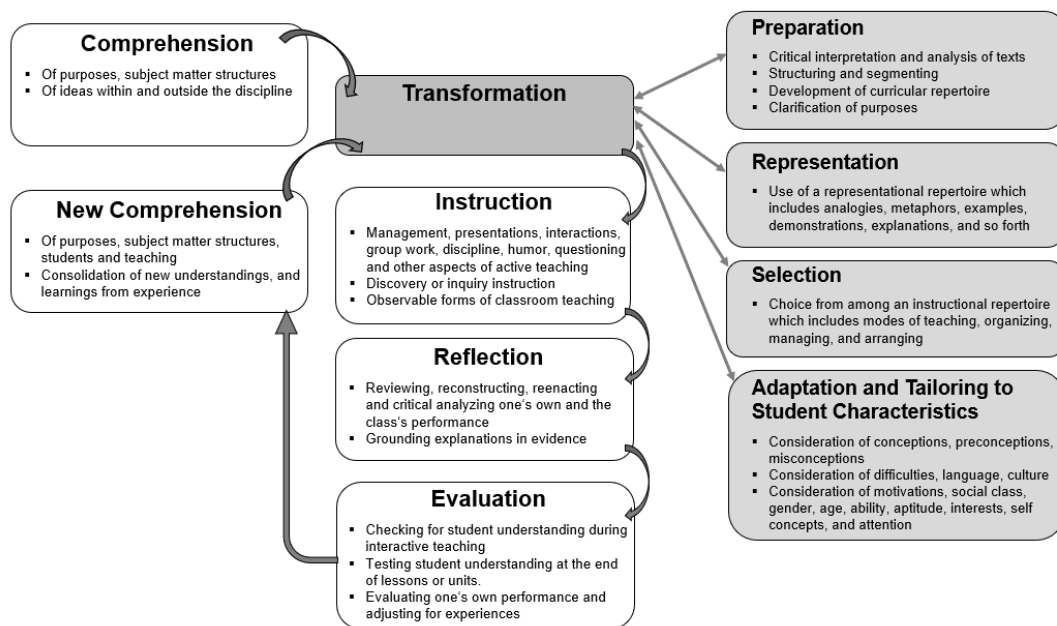


Abbildung 25: *Model of pedagogical reasoning and action (MPRA)*, (Shulman, 1987, S. 15 sowie Fernandez, 2014, S. 82, angepasst durch Autorin)

Transformationsmodelle wie dieses erweisen sich gemäss Kind (2009) gerade für die Ausbildung angehender Naturwissenschaftslehrpersonen als nützlich, denn sie implizieren, dass die Entwicklung von PCK einem Mechanismus unterliegt, der auch eine gute theoretische Grundlage für die Ausbildung darstellen könnte. Shulman hat bereits 1987 einen solchen Mechanismus in seinem „*model of pedagogical reasoning and action*“ (MPRA) skizziert (Shulman, 1987, S. 15; Abb. 25). Das fachdidaktische Wissen wird in diesem Modell als eigenständige Wissensdimension aufgefasst, die unter anderem Wissen darüber beinhaltet, wie bestimmte Sachverhalte in Bezug auf Lernprozesse, Schülerfähigkeiten und -interessen organisiert und dargestellt werden können, damit das Lernen von Schülerinnen und Schülern unterstützt wird (Shulman, 1987, cf. Kirschner et al., 2017). Dieses dynamische Modell von Shulman (1987) verdeutlicht die Entwicklung von PCK von Lehrpersonen und unterstreicht die Relevanz der Reflexion in diesem Prozess (Fernandez, 2014, S. 81). Das Modell weist in gewisser Weise auch Ähnlichkeiten zum Planungsmodell der «Didaktischen Rekonstruktion» auf, in dem das fachdidaktische Wissen bei der Planung von Unterricht zur Anwendung kommt. Allerdings wird im Modell der «Didaktischen Rekonstruktion» die Reflexion nicht explizit mitberücksichtigt, dafür der Themenbezug ganz bewusst verlangt (Kattmann, Duit, Gropengiesser & Komorek, 1997; cf. Kap. I, 2.8.4).

Für die iterative Weiterentwicklung von Unterrichtsplanungen und für die Festigung der fachdidaktischen Handlungspläne stellen Reflexionsanlässe demnach eine wichtige Voraussetzung dar (Stender, 2014, S. 39; Stender et al., 2017). Dies konnten van der Valk &

Broekman (1999) in ihrer Studie bestätigen. Durch die Planung von Unterricht zu Themen aus den Fächern Physik und Mathematik und einem anschliessenden Gespräch darüber haben angehende Lehrpersonen nicht nur ihr vorhandenes PCK gezeigt, sondern auch weiterentwickelt. Auch in einer aktuellen Studie von Anthofer zum Planen und Unterrichten des Themas «Experimentieren bei Stofftrennungen» konnte eine Zunahme des fachdidaktischen Wissens und Fachwissens von angehenden Lehrpersonen festgestellt werden, was ausserdem mit einer Steigerung der Unterrichtsqualität verbunden war (Anthofer, 2017, S. 197). Gemäss dem Autor könnten hier Parameter wie u.a. der kollegiale Austausch oder aber die Analyse der Videos eine förderliche Wirkung gezeigt haben. Ozogul, Olina, & Sullivan (2008), die ein Kodiermanual zur Erfassung geplanter Lernprozesse zum Thema Technik in Unterrichtsplanungen entwickelt haben, stellten fest, dass die Anwendung dieses Manuals in verschiedenen Evaluations- und Feedback-Settings (Feedback durch Lehrperson, Feedback von Peers, Selbstevaluation und -reflexion) insgesamt zu einer Qualitätsverbesserung der Planungen geführt hat. Eine genauere Analyse hat aber auch gezeigt, dass aufgrund des Lehrerfeedbacks ein höherer Qualitätszuwachs zu verzeichnen war. In diesem Zusammenhang stellten Jacobs, Martin & Otieno (2008) in ihrer Untersuchung fest, dass Resultate aus der Evaluation von Unterrichtsplanungen zu Gesprächsanlässen führen, die angehende Lehrpersonen in ihrer fachdidaktischen Entwicklung unterstützen können. Die hier dargelegten Befunde deuten darauf hin, dass die Planungskompetenzen mit Hilfe unterschiedlicher Reflexionsanlässe gefördert worden sind.

In der Lehrpersonenbildung haben sich mittlerweile verschiedene Formen von Begleitung und Unterstützung etabliert, mit deren Hilfe angehende Lehrpersonen in ihrer Kompetenzentwicklung unterstützt werden können (Schnebel, Kreis & Musow, 2017). Konzepte des Peer-Coachings stellen hierzu einen möglichen und vielversprechenden Ansatz zur Anreicherung praxisbezogener Lerngelegenheiten dar (Britton & Anderson, 2010; Schnebel et al., 2017). Dabei sind Peer-Coaching-Settings reziprok angelegt. Sie zielen darauf ab, dass sich angehende Lehrpersonen mehr oder weniger systematisch angeleitet gegenseitig in der Bearbeitung studiumsbezogener Aufgaben und damit in ihrer Kompetenzentwicklung unterstützen (cf. Kreis & Schnebel, 2017). Das kollegiale Unterrichtscoaching (KUC) repräsentiert dabei eine mögliche Form von Peer-Coaching. Im Zentrum dieses Ansatzes steht eine relativ kurze Fortbildung, in der sich die sich coachenden Lehrpersonen, die formal gleichrangig sind, auf der Basis von sogenannte Kernaspekten zur Planung und Reflexion von Unterricht ihr Handlungsrepertoire aneignen, das zielorientierte und vertiefte unterrichtsbezogene Kooperationsprozesse und Innovation begünstigt (Kreis & Staub, 2009). Ziel des Coachings ist die optimale Ausrichtung unterrichtlichen Handelns auf das Lernen der Schülerinnen und Schüler, wobei der Fokus konsequent auf fachspezifische Lehr- und

Lernprozesse gelegt wird (ebd.). Studien zu praxissituierten Lernprozessen zeigen insgesamt eine Intensivierung des Lernertrages, sofern sich die Beteiligten an Coaching- oder Mentoringmodellen beteiligen (Futter, 2017; Kreis & Schnebel, 2017; Kreis & Staub, 2011). Die unterstützenden Akteurinnen und Akteure waren bisher allerdings vor allem Dozierende oder Lehrpersonen (Kreis & Schnebel, 2017). Es ist aber zu erwarten und konnte bereits empirisch belegt werden, dass auch systematisch-kooperative Formen des Lehrens und Lernens zwischen Peers mit ähnlichen Bedürfnissen und ähnlichem Erfahrungshintergrund für deren Kompetenzentwicklung von Bedeutung sind (Kreis & Schnebel, 2017; Schnebel et al., 2017). Im Projekt KUBeX wurde daher im Rahmen der Interventionsstudie die Wirkung des Content-Focused Peer Coaching-Ansatzes auf das Planungshandeln angehender Lehrpersonen der Sekundarstufe I untersucht (cf. Kap. I, 3.1). Der erste Teil des dafür durchgeführten dyadischen Planungsgesprächs, in welchem die Studierenden sich jeweils gegenseitig ihre Planungen vorstellen mussten, wurde auch für die vorliegende Studie und die damit verbundene Analyse der Qualität der Unterrichtsplanungen zum Experimentieren im Sinne der Erkenntnisgewinnung hinzugezogen. Porter (2002) sieht in der Aufgabe, Unterrichtsplanungen möglichst genau beschreiben zu müssen, eine Möglichkeit, wie Lehrpersonen ihre Reflexionsfähigkeit und damit auch ihren Unterricht verbessern können. Aus Sicht der Forschung bieten solche Anlässe ausserdem die Möglichkeit, weitere Beobachtungen zu subjektiv geprägten, aber unterrichtsrelevanten Lernereignissen in Form von Äusserungen oder Gesten zu generieren (cf. Irion, 2010; Kreis & Staub, 2012). Bevor in Kapitel II, 1.5 die Teilstudie II der vorliegenden Studie näher vorgestellt wird, fasst Kapitel II, 1.4 den bis hier dargestellten theoretischen und empirischen Hintergrund, auf dem die Teilstudie II beruht, im Sinne eines Fazits zusammen.

1.4 Fazit

Für die Gestaltung von qualitativem Unterricht nimmt die Vorbereitung (Planung) und Nachbereitung (Reflexion) eine wichtige Rolle ein, insbesondere bei angehenden Lehrpersonen (Staub & Kreis, 2013; Stender et al., 2015). Das Planungshandeln stellt daher ein wichtiges Element unterrichtlicher Praxis dar und bietet einen geeigneten Anlass, das fachdidaktische Wissen in einer handlungsnahen Situation anzuwenden (Smith & Banilower, 2015) und aufzubauen (Brown, Friedrichsen & Abell, 2013). Mit den neuen Bildungsstandards (EDK, 2011; KMK, 2005; NCR, 2012) sind (angehende) Lehrpersonen verpflichtet, Kompetenzen im Bereich des Experimentierens im Sinne der Erkenntnisgewinnung zu fördern. Damit verbunden ist auch der Aufbau eines Verständnisses derjenigen Prozesse, mit denen naturwissenschaftliche Erkenntnisse generiert werden und die durch Konventionen und ethische Grundsätze gekennzeichnet sind (Roberts, 2007; Schwartz et al., 2004). Für

die Gestaltung eines solch komplexen Unterrichts sollten (angehende) Lehrpersonen über ein fundiertes themenspezifisches Professionswissen und die damit verbundenen professionellen Handlungskompetenzen verfügen (Crawford, 2000, 2007), insbesondere auch im Biologieunterricht. Entsprechend sollte sich dieses Wissen in der Qualität des geplanten Unterrichts manifestieren. Allerdings zeigen diverse Studien, dass naturwissenschaftliche Arbeitsweisen wie das Experimentieren im Sinne der Erkenntnisgewinnung im Unterricht selten im Sinne eines Fachinhaltes vermittelt werden (e.g. Capps & Crawford, 2013a; Schulz, Riess, Wirtz, & Barzel, 2012). Naturwissenschaftliche Arbeitsweisen werden in der Regel dazu eingesetzt, Schülerinnen und Schüler für Naturwissenschaften zu motivieren, Phänomene darzustellen, fachliche Inhalte und Konzepte zu erarbeiten oder das praktische Arbeiten zu fördern (e.g. Gyllenpalm & Wickman, 2011a; Tesch & Duit, 2004). Die Vermutung, dass (angehende) Lehrpersonen hinsichtlich des Lehrens von Kompetenzen wissenschaftlichen Denkens und Arbeitens Defizite aufweisen, konnte durch diverse Studien gestützt werden (e.g. Chabalengula, Mumba & Mbewe, 2012; Davis, Petish, & Smithy, 2006; Gyllenpalm & Wickman, 2011a; Schneider & Plasman, 2011; Crawford, 2007). Diese haben gezeigt, dass es (angehenden) Lehrpersonen u.a. an fachlichem Wissen und Verständnis zum Konzept *Scientific Inquiry*, an fachdidaktischem Wissen einen entsprechenden Unterricht zu gestalten, aber auch an Überzeugungen zur Lehrtätigkeit und zum Prozess der Erkenntnisgewinnung mangelt (cf. Roehrig & Luft, 2004). Vor diesem Hintergrund kommt dem Professionalisierungsprozess angehender Lehrpersonen eine bedeutende Rolle zu (Wilson, Taylor, Kowalski & Carlson, 2010). Es bedarf geeigneter Anlässe, mit deren Hilfe das notwendige fachdidaktische Wissen aufgebaut werden kann. Allgemein wird davon ausgegangen, dass die Analyse von schriftlichen Unterrichtsplanungen Aussagen über die Qualität von Unterrichtskonzeptionen als Antizipation einer Handlungssituation ermöglichen (e.g. König et al., 2015). Dabei scheinen Planungsraster im Gegensatz zu Planungsmodellen bei Berufsanfängerinnen und Berufsanfängern stärker einzuschränken und zu weniger Planungsüberlegungen zu führen (Zierer, Werner & Wernke, 2015). Operationalisierungen der Planungskompetenz sind allerdings bisweilen ganz allgemein, aber auch hinsichtlich der Kompetenzstufungen, noch ungeklärt (Zierer, Werner, et al., 2015, S. 378) und es gibt entsprechend auch nur wenige Studien mit einem naturwissenschaftsdidaktischen Fokus, die Aussagen zur Planungskompetenz und damit zur Qualität der Unterrichtsplanungen erlauben (Stender, 2014; Stender et al., 2017). Diese Studien fokussieren aber Planungen zu einem fachlichen und nicht zu einem fachmethodischen Thema. Entsprechend ist wenig erforscht, ob es (angehenden) Lehrpersonen gelingt, Unterricht zu planen und zu analysieren, der auf den Aufbau von Kompetenzen naturwissenschaftlicher Erkenntnisgewinnung abzielt (Schneider & Plasman, 2011; cf. auch Kap. II 1.3). Auch ist wenig bekannt, welche Qualität unterrichtliche Planungen von Studierenden

zum Experimentieren im Sinne der Erkenntnisgewinnung erreichen und durch welche Faktoren (Fachwissen und fachdidaktisches Wissen zum Unterrichten naturwissenschaftlicher Erkenntnisgewinnung) die Planungsqualität beeinflusst wird.

Teilstudie II des Dissertationsvorhabens greift diese Lücke auf und untersucht vor diesem Hintergrund die im Rahmen des Projekts KUBeX von angehenden Lehrpersonen gestalteten Unterrichtsplanungen, -materialien und videografierten Planungsgespräche in Bezug auf das Vorkommen und die Qualität experimentellen Handelns im Sinne der Erkenntnisgewinnung. Auch werden mögliche Zusammenhänge zwischen der ermittelten Planungsqualität und dem im Projekt KUBeX getesteten Professionswissen ermittelt. Im Sinne eines Überblicks wird Teilstudie II im nächsten Kapitel vorgestellt.

1.5 Überblick über Teilstudie II

Im Zentrum der Teilstudie II steht die Analyse der Qualität von Unterrichtsplanungen, -materialien und videografierten Planungsgespräche, die im Projekt KUBeX entstanden sind. Die Analyse dieses zweiten Forschungsanliegens erfolgt explorativ mit Hilfe des im Projekt KUBeX entwickelten Ratingmanuals, das in Teilstudie I auf der Basis der erhobenen Daten inhaltlich und faktoriell validiert werden konnte (Kap. I, 4.2). In einem ersten Schritt werden die Unterrichtsplanungen, -materialien und videografierten Planungsgespräche auf Hinweise analysiert, die auf einen Unterricht zum Experimentieren im Sinne der Erkenntnisgewinnung schliessen lassen. Planungsentscheidungen und das in diesem Zusammenhang beobachtbare fachdidaktische Wissen werden auf der Basis der Gesamtstichprobe ($N = 119$) quantifiziert. In einem nächsten Schritt werden diejenigen Unterrichtsplanungen, -materialien und videografierten Planungsgespräche der Teilstichprobe ($n = 70$), in denen das fachdidaktische Wissen zur Planung von Unterricht zum experimentellen Handeln zumindest ansatzweise beobachtet werden konnte, mit Hilfe des im Projekt eingesetzten Planungsrasters hinsichtlich der Qualität der Umsetzung ausgewertet. Geprüft werden auch Auswirkungen auf die beiden Dimensionen des Modells der «Didaktischen Rekonstruktion», «Analyse» und «Konstruktion», die bei der Validierung des Instruments ermittelt werden konnten (cf. Kap. I, 2.8.4). Wie gut die im Projekt KUBeX untersuchten Gruppen (Interventions- und Kontrollgruppe) den Auftrag, eine Doppellektion zum Thema «Experimentieren-lernen am Themenbereich «Visuelle Wahrnehmung» (Gropengiesser, 1997, 2001, 2007)» umgesetzt haben, soll anschliessend noch untersucht werden.

Die Ergebnisse der Qualitätsanalyse der Teilstichprobe ($n = 70$) werden zum Schluss in Bezug zu den Ergebnissen der Professionswissenstests gesetzt, die im Rahmen des Projekts KUBeX entwickelt und durchgeführt worden sind. Damit sollen mögliche Zusammen-

hänge zwischen dem Testwissen (Disposition) und dem in den Planungen erfassten, handlungsnäheren fachdidaktischen Wissen (Performanz) geprüft werden.

Vor diesem Hintergrund wurden entsprechende Forschungsfragen ausformuliert. Diese werden im nächsten Kapitel (Kap. II, 1.6) näher vorgestellt, zusammen mit den Hypothesen, die von den dargelegten theoretischen und empirischen Befunden abgeleitet worden sind.

1.6 Forschungsfragen und Hypothesen

Teilstudie II der vorliegenden Dissertation verfolgt insgesamt zwei Ziele. Im Rahmen des ersten Ziels sollen die Unterrichtsplanungen, -materialien und videografierten Planungsgespräche im Hinblick auf das Vorkommen und die Qualität experimentellen Handelns untersucht werden (Forschungsanliegen 2a). Das zweite Ziel befasst sich mit der Frage nach möglichen Zusammenhängen zwischen der erfassten Qualität der Planungen und weiteren, im Projekt KUBeX untersuchten Faktoren, wie dem getesteten Fachwissen und dem fachdidaktischen Wissen (Forschungsanliegen 2b). Die Forschungsfragen sind daher entsprechend in zwei Bereiche eingeteilt und werden in der Folge präzisiert.

I. Vorkommen und Qualität experimentellen Handelns in Planungen (Forschungsanliegen 2a)

Die Förderung von Kompetenzen im Bereich der Erkenntnisgewinnung wird, obwohl in den Bildungsstandards verankert (EDK, 2011; KMK, 2005; NCR, 2012), wie in Kap. II, 1.2 dargelegt, aufgrund fachlicher und fachdidaktischer Schwierigkeiten, aber auch fehlender Überzeugungen in der Praxis wenig umgesetzt (e.g. Capps & Crawford, 2013b; Crawford, 2007; Davis et al., 2006; Gomez-Zwiep, 2008; Gyllenpalm & Wickman, 2011b). Diese empirischen Befunde beruhen u.a. auf Interviews. Bisher wurden keine Qualitätsanalysen von Planungen zum experimentellen Handeln im Sinne der Erkenntnisgewinnung auf der Basis eines Ratingmanuals durchgeführt (cf. Kap. I, 2.8.3). Vor dem Hintergrund, dass alle am Projekt KUBeX beteiligten angehenden Lehrpersonen sich im Rahmen der Impulsveranstaltung I 2 \times 90 Minuten mit fachlichen und fachdidaktischen Aspekten zum experimentellen Handeln und zum Prozess der Erkenntnisgewinnung auseinandergesetzt haben, wird angenommen, dass diese motiviert waren, eine Doppellektion zum Experimentieren im Sinne der Erkenntnisgewinnung zu planen und dass sich gemäss Transformationsmodell (cf. Kap. I, 2.8.2) ihr fachdidaktisches Wissen in den Planungen zeigt. Allerdings ist es wahrscheinlich, dass die Qualität der Umsetzungen vom verwendeten Planungsraster beeinflusst wird (Zierer, Werner, et al., 2015). Unterschiede zwischen der

Interventions- und der Kontrollgruppe sind zum Messzeitpunkt aufgrund der vorangehend erfolgten Intervention zum Peer-Coaching in der Interventionsgruppe ausserdem wahrscheinlich und müssten sich daher in einer qualitativ besseren Planung zeigen.

Vor diesem Hintergrund lauten die Forschungsfragen mit den dazu formulierten Hypothesen:

1. Setzen angehende Lehrpersonen der Sekundarstufe I das experimentelle Handeln in ihren Unterrichtsplanungen im Sinne der Erkenntnisgewinnung um?

H1: Angehende Lehrpersonen, die die Impulsveranstaltung I zum experimentellen Handeln besucht haben und einen entsprechenden Unterricht am Thema «Visuelle Wahrnehmung» planen müssen, gestalten diesen grösstenteils mit dem Fokus auf dem experimentellen Handeln als Fachinhalt.

2. In welcher Qualität planen angehende Lehrpersonen mit Hilfe des im Projekt verwendeten Planungsrasters einen Unterricht zum experimentellen Handeln im Sinne naturwissenschaftlichen Erkenntnisgewinns?

H2: Wenn angehende Lehrpersonen auf der Basis des vorgegebenen Planungsrasters einen Unterricht zum experimentellen Handeln planen, dann wird die Qualität der Planungsüberlegungen durch das Planungsinstrument beeinflusst.

3. Unterscheiden sich die im Projekt KUBeX untersuchten Interventions- und Kontrollgruppen hinsichtlich der Qualität ihrer Unterrichtsplanungen zum experimentellen Handeln?

H3: Die Interventionsgruppe plant aufgrund der in einer Intervention erfolgten, vertieften Auseinandersetzung mit den Kernaspekten einen qualitativ besseren Unterricht zum experimentellen Handeln als die Kontrollgruppe.

II. Zusammenhänge zwischen der Planungsqualität und den Ergebnissen der Professionswissenstests (Forschungsanliegen 2b)

Empirische Befunde zu möglichen Zusammenhängen zwischen dem Professionswissen von (angehenden) Lehrpersonen und den Schülerleistungen sind bisher nicht eindeutig (cf. Kap. II, 1.1.1). Es gibt allerdings Hinweise, wonach das Niveau des Fachwissens einen mediierenden Einfluss darauf hat, wie der Fachinhalt unterrichtet wird (Carlsen, 1991; Gess-Newsome & Lederman, 1995) und dass das Fachwissen eine notwendige, aber nicht hinreichende Voraussetzung für die Gestaltung eines qualitätsvollen Unterrichts darstellt (u.a. Kunter et al. 2011). Demgegenüber scheint das deklarative fachdidaktische Wissen einer Lehrperson ein guter Prädiktor für die Qualität von Schülerleistungen, aber auch für die kognitive Struktur der im Unterricht verwendeten Lerngelegenheiten zu sein (Baumert et al., 2010; Hill et al., 2005; Olszewski, 2010). Ungeklärt ist aber, ob sich diese Ergebnisse auf den Biologieunterricht übertragen lassen. Daher lauten die hierzu formulierten Forschungsfragen:

4. Lassen sich Zusammenhänge zwischen dem deklarativen fachdidaktischen Wissen und der Qualität der Unterrichtsplanungen, -materialien und videografierten Planungsgespräche in Bezug auf das experimentelle Handeln im Biologieunterricht feststellen?

H4: Wenn angehende Lehrpersonen im Test ein gutes fachdidaktisches Wissen zeigen, dann planen sie einen qualitativ besseren Unterricht zum experimentellen Handeln als Fachinhalt.

5. Lassen sich Zusammenhänge zwischen dem deklarativen Fachwissen und der Qualität der Unterrichtsplanungen, -materialien und videografierten Planungsgespräche in Bezug auf das experimentelle Handeln im Biologieunterricht feststellen?

H5: Wenn angehende Lehrpersonen im Test ein gutes Fachwissen zeigen, dann planen sie einen qualitativ besseren Unterricht zum experimentellen Handeln als Fachinhalt.

2 Forschungsdesign und Methode

Auf der Basis der im vorangehenden Kapitel formulierten Forschungsfragen und Hypothesen werden in diesem Kapitel das Forschungsdesign und die Methoden der Teilstudie II vorgestellt. Dabei wird auf eine erneute Beschreibung der Stichprobe und des Ratingverfahrens verzichtet, da diese beiden Aspekte bereits in Teilstudie I näher erläutert worden sind (cf. Kap. I, 3.2.2 & I, 3.2.4). Im Zentrum dieses Kapitels stehen daher nur die Beschreibung der Datengrundlage, auf der die Beantwortung der Forschungsfragen beruht (cf. Kap. II, 2.1), sowie die Präsentation des Professionswissenstests (Kap. II, 2.2), der im Rahmen des Projekts KUBeX durchgeführt worden ist. Letzterer soll, ausgehend von den vorhandenen Unterlagen, vorgestellt werden.

2.1 Design

Wie in Teilstudie I ermittelt, hat sich das theoriebasiert entwickelte Erhebungsinstrument zur Erfassung des fachdidaktischen Wissens zur Planung von Unterricht zum experimentellen Handeln als valide herausgestellt (Kap. I, 4.1). Ein erstes Ziel der Teilstudie II ist daher, eine Analyse der Unterrichtsplanungen, -materialien und videografierten Planungsgespräche mit Hilfe dieses Instruments im Hinblick auf das Vorkommen und die Qualität experimentellen Handelns vorzunehmen. Im Rahmen des zweiten Zieles werden die dabei gewonnenen Erkenntnisse in Beziehung zu den Ergebnissen der Onlinefragebogen gesetzt. Das auf diese beiden Zielsetzungen ausgerichtete Design der Teilstudie II ist in Abbildung 26, eingebettet in das Gesamtprojekt KUBeX, dargestellt. Die für die Teilstudie II wesentlichen Datenquellen und Erhebungsverfahren sind auch hier in zwei verschiedenen Grautönen speziell hervorgehoben. Die Datengrundlage und das Ratingverfahren zur Erfassung des fachdidaktischen Wissens zur Planung von Unterricht zum experimentellen Handeln wurden bereits in Kap. I, 3.2.4 näher beschrieben. Für die Ermittlung möglicher Zusammenhänge zwischen der Qualität der Unterrichtsplanungen zum experimentellen Handeln (Performanz) und dem deklarativen, themenspezifischen Professionswissen (Disposition) wurden im Projekt KUBeX darüber hinaus Professionswissenstests entwickelt und durchgeführt. Dabei wurden alle 119 Studierenden der vier Pädagogischen Hochschulen zu drei verschiedenen Messzeitpunkten (Prä-, Post-, Postpostdesign) getestet (Abb. 26).

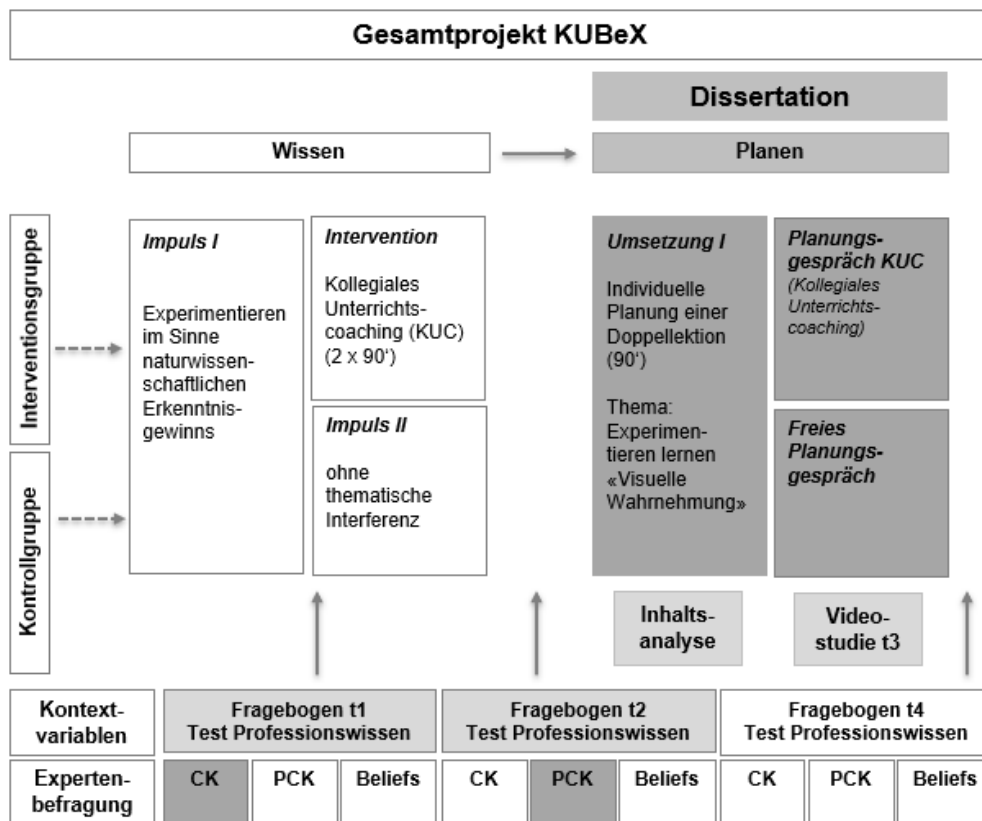


Abbildung 26: Design der Teilstudie II eingebettet in das Gesamtdesign des Projekts KUBeX; Teilstudie II (grau hinterlegt): Datenquellen (dunkelgrau), Erhebungsverfahren (hellgrau)

In Anlehnung an das für das Projekt wegleitende Modell der COACTIV-Studie (Baumert & Kunter, 2006, S. 482, cf. Kap. I, 2.4.3) und die darin enthaltenen Kategorien zum Professionswissen von Shulman (1986/87) wurden das deklarative Fachwissen (*content knowledge*), das deklarative fachdidaktische Wissen (*pedagogical content knowledge*) sowie die Beliefs mit Hilfe von Onlinefragebogen erhoben. Das pädagogische Wissen, bei dem angenommen werden kann, dass es im Gegensatz zu den anderen Wissensarten keine fachimmanenten Anteile enthält und deswegen nur vermittelt über das fachdidaktische Wissen Einfluss auf die Handlungsskripte des jeweiligen Faches nimmt, wurde aus testökonomischer Sicht nicht geprüft (cf. Stender, 2014, S.125). Die Testinstrumente zur Erfassung des deklarativen fachdidaktischen Wissens und des deklarativen Fachwissens werden im nächsten Kapitel (Kap. II, 2.2) vorgestellt. Für Untersuchungen zu den Beliefs sei auf Arbeiten von Smit, Rietz und Kreis (2017) und Smit, Weitzel et al. (2017) verwiesen.

2.2 Professionswissenstest

Zur Erfassung des themenspezifischen Professionswissens wurde im Rahmen des Projekts KUBeX ein Professionswissenstest entwickelt. Daraus sollen hier die beiden für die Teilstudie II relevanten Tests vorgestellt werden. Die Entwicklung und Auswertung der Tests ist Gegenstand eines zweiten Dissertationsvorhabens, von dem Ausschnitte an der ERIDOB (*european researchers in didactics of biology*) in Karlstad vorgetragen worden sind (Blank, 2016). Die folgenden Ausführungen zu den beiden Wissenstests (deklaratives fachdidaktisches Wissen und deklaratives Fachwissen) stützen sich einerseits auf die Angaben aus dem Vortrag, andererseits aber auch auf Beschreibungen aus Arbeiten von Smit, Rietz et al. (2017) und Smit, Weitzel et al. (2017).

Die für die Teilstudie II im Zentrum stehenden Dimensionen des Professionswissens wurden mit Hilfe neu entwickelter Items erfasst, die ausgerichtet auf die zentralen Inhalte im Projekt waren. Dabei wurden insgesamt 21 Items für den Fachwissenstest und 22 Items für die Erhebung des fachdidaktischen Wissens entwickelt. Die Fachwissensitems wurden zum einen auf die Erfassung des biologischen Fachinhalts der visuellen Wahrnehmung ausgerichtet, zum anderen aber auch auf den fachmethodischen Fachinhalt zum Prozess der Erkenntnisgewinnung. Die Entwicklung der Items zur visuellen Wahrnehmung erfolgte auf der Basis eines Forschungsprojekts von Dannemann & Krüger (2010), die im Hinblick auf die spätere Diagnose von Schülervorstellungen Items zur Erfassung der visuellen Wahrnehmung entwickelt haben. Diese Items waren gemäss Blank (2016) reliabel und wurden in die Testung aufgenommen. Demgegenüber waren die Items zum fachmethodischen Fachinhalt, dem experimentellen Handeln, nicht reliabel (ebd.). Sie konnten daher für die Tests nicht eingesetzt werden. Die Items zum fachdidaktischen Wissen basierten auf den beiden Dimensionen «Kenntnis zu den Schülervorstellungen» und «Kenntnis zu Instruktionsstrategien zum Experimentieren», zwei Wissensfacetten, die in der naturwissenschafts-didaktischen Forschung als besonders relevant eingestuft werden. Sie wurden gemäss den drei von Jüttner & Neuhaus (2012) vorgeschlagenen Wissensdimensionen (deklarativ, prozedural und konditional) aufgebaut (Kap. I, 2.5.2). Insgesamt mussten von den Studierenden drei Tests online gelöst werden, wobei t1 am eigenen Computer in einer Modulveranstaltung bearbeitet werden konnte, während t2 und t4 jeweils innerhalb einer Woche am eigenen Computer durchgeführt werden mussten.

Für die Analyse möglicher Zusammenhänge zwischen den beiden Wissenstests und der Qualität der Unterrichtsplanungen, -materialien und videografierten Planungsgespräche wurden in der Teilstudie II die Ergebnisse aus t1 zum deklarativen Fachwissen zur visuellen Wahrnehmung und die Ergebnisse aus t2 zum deklarativen fachdidaktischen Wissen

zur Planung eines Unterrichts zum experimentellen Handeln beigezogen. Auf den Fachwissenstest zum experimentellen Handeln (t2) musste aus oben erwähnten Gründen verzichtet werden. Die Ergebnisse zu den jeweiligen Forschungsfragen werden in Kapitel II, 3 entlang der Forschungsfragen dargestellt. Die dafür beigezogenen statistischen Verfahren werden zunächst aber noch im nächsten Kapitel (Kap. II, 2.3) vorgestellt.

2.3 Statistische Verfahren

Zur Beantwortung der Forschungsfragen zum Vorkommen und zur Qualität fachdidaktischen Wissens in Planungen von Unterricht zum experimentellen Handeln wurden alle Analysen in SPSS (IBM Corp., 2016) durchgeführt. Dabei wurde in einem ersten Schritt eine deskriptive, univariate Analyse vorgenommen, um Summenscores und damit verbundene Häufigkeitsverteilungen zu berechnen. Auf dieser Grundlage wurde auch die Kompetenzausprägung der Studierenden in Bezug auf die Gestaltung von Unterricht zum experimentellen Handeln ermittelt und eingeordnet. Für die weiteren inferenzstatistischen Analysen mussten im Vorfeld die statistischen Voraussetzungen wie Normalverteilung und Varianzhomogenität mittels Shapiro-Wilk-Test (Test auf Normalverteilung) und Levene-Test (Test auf Varianzhomogenität) geprüft werden (Döring & Bortz, 2016, S. 617). Die Berechnung von Mittelwerten erfordert prinzipiell ein metrisches Skalenniveau. In der Forschungspraxis werden Ratingskalen aber häufig auch als intervallskaliert aufgefasst (Döring & Bortz, 2016, S. 251). Voraussetzung dafür ist, dass die Antworten gleichabständig formuliert sind und mindestens fünf Stufen aufweisen (ebd.). Ersteres wurde in der vorliegenden Studie normativ angelegt, letzteres ist allerdings nicht ganz erfüllt. Daher sind alle Mittelwertvergleiche und Korrelationen zwischen den Kategorien bzw. Dimensionen sowohl parametrisch als auch nichtparametrisch durchgeführt worden. Für die statistisch vergleichenden Analysen der Mittelwerte der Teilstichprobe ($n = 70$) und der Reststichprobe ($n = 49$), aber auch der Interventions- und der Kontrollgruppe wurden der t-Test und der Mann-Whitney-U-Test für unabhängige Stichproben verwendet (Field, 2013). Die Mittelwerte für die im Ratingmanual normativ festgelegten 5 Kategorien, aber auch für die in Teilstudie I festgestellten Dimensionen «Analyse» und «Konstruktion» wurden auf der Basis der Teilstichprobe ($n = 70$) mit Hilfe des t-Tests als auch nichtparametrisch mit dem Wilcoxon-Test für verbundene Stichproben verglichen (Field, 2013). Da mit der Festlegung des Signifikanzniveaus keine Aussagen zur Grösse der erfassten Unterschiede gemacht werden können, wurden zur Ermittlung der Unterschiede die Effektstärken in Form des Korrelationskoeffizienten r erhoben (Bühner, 2011). Die Effektstärke stellt eine standardisierte Messung dar, die Vergleiche ermöglicht. Werte von .10 werden als schwacher Effekt eingestuft, da damit nur 1% der Gesamtvarianz erklärt wird. Werte von .30 werden

als mittlerer Effekt interpretiert und Werte von .50, die 25% der Gesamtvarianz erklären, werden als starke Effekte aufgefasst (Cohen, 1992; cf. Field, 2013, S. 82). Die bivariaten Korrelationen der Kategorien bzw. Dimensionen wurden ebenfalls parametrisch (Pearson-Korrelation) als auch nichtparametrisch (Spearman-Korrelation) ausgewertet (Field, 2013, S. 274f.). Die Korrelationen der Kategorien wurden aufgrund der Alpha-Fehler-Kumulation mit Hilfe der Bonferroni-Korrektur bereinigt (Schendera, 2014). Die Einordnung der Interkorrelationen der Mittelwerte der Kategorien bzw. der Dimensionen erfolgte auf der Basis der Wertetabelle von Weiber & Mühlhaus (2014, S. 15), für die Korrelationen zwischen den Testergebnissen und der Qualität der Planungsmaterialien wurde die Einteilung nach Cohen 1992 (Field, 2013, S. 82) verwendet.

Zur Beantwortung der Forschungsfragen zu möglichen Zusammenhängen zwischen den Testergebnissen aus den Professionswissenstests und den Ergebnissen aus den Ratings der Unterrichtsplanungen, Arbeitsmaterialien und Planungsgesprächen wurden die Analysen in SPSS (IBM Corp., 2016) durchgeführt. Es wurden bivariate Korrelationen sowohl parametrisch wie auch nichtparametrisch berechnet (siehe oben). Mit den errechneten Korrelationen zwischen der ermittelten Qualität des fachdidaktischen Wissens zur Planung von Unterricht zum experimentellen Handeln und den Ergebnissen aus den Professionswissenstests soll ausserdem die faktorielle Validität des Ratinginstruments nochmals aus einer anderen Perspektive geprüft werden.

3 Ergebnisse

Die Struktur des vorliegenden Kapitels wird durch die beiden in Kapitel II, 1.6 vorgestellten Ziele, die dazu formulierten Forschungsfragen und die in diesem Zusammenhang erfolgten Untersuchungen vorgezeichnet.

Ausgehend von der ersten Zielsetzung (Forschungsanliegen 2a) werden zunächst die empirischen Befunde zum Vorkommen und zur Qualität experimentellen Handelns in den Unterrichtsplanungen, -materialien und videografierten Planungsgesprächen präsentiert (Kap. II, 3.1). Dabei werden in einem ersten Schritt Ergebnisse zum Vorkommen experimentellen Handelns in den Unterrichtsplanungen dargestellt und erläutert (Forschungsfrage 1; cf. Kap. II, 3.1.1). Diese beruhen auf der Gesamtstichprobengrösse von $N = 119$. Erkenntnisse zur Qualität der Unterrichtsplanungen, -materialien und videografierten Planungsgespräche werden in einem zweiten Schritt vor dem Hintergrund des im Projekt eingesetzten Planungsrasters und mit Bezug zu den beiden Dimensionen des Modells der «Didaktischen Rekonstruktion», «Analyse» und «Konstruktion», beschrieben (Forschungsfrage 2; cf. Kap. II, 3.1.2). Sie beruhen auf der Teilstichprobe von $n = 70$, die das experimentelle Handeln in ihren Unterrichtsplanungen, -materialien und videografierten Planungsgesprächen zumindest ansatzweise umgesetzt haben. Ergebnisse zur Analyse möglicher Unterschiede zwischen Interventions- und Kontrollgruppe hinsichtlich der qualitativen Umsetzung der gestalteten Unterrichtsplanungen werden in einem dritten Schritt in Kapitel II, 3.1.3 dargelegt (Forschungsfrage 3). Die Präsentation der Ergebnisse erfolgt bei Nichtvorliegen von Normalverteilung auf der Basis der nichtparametrisch ermittelten Befunde. Die Ergebnisse aus den parametrisch erfolgten Analysen sind, sofern sinnvoll, jeweils im Anhang einzusehen.

Im Rahmen der zweiten Zielsetzung (Forschungsanliegen 2b) werden anschliessend die empirischen Befunde zu möglichen Zusammenhängen zwischen der Qualität der Unterrichtsplanungen, -materialien und videografierten Planungsgespräche und den Professionswissenstests präsentiert (Kap. II, 3.2). Dabei werden zunächst die Erkenntnisse zum Einfluss des deklarativen fachdidaktischen Wissens auf die Planungsqualität beschrieben (Forschungsfrage 4; cf. Kap. II, 3.2.1). Das Ergebniskapitel schliesst mit den Ergebnissen der Analyse eines möglichen Zusammenhangs zwischen dem deklarativen Fachwissen und der Planungsqualität (Forschungsfrage 5; cf. Kapitel II, 3.2.2).

3.1 Vorkommen und Qualität von Unterrichtsplanungen zum experimentellen Handeln



Die bisherigen Analysen des im Projekt KUBeX entwickelten Ratingmanuals haben gezeigt, dass das Instrument auf der Basis der erhobenen Daten faktoriell valide ist (Kap. I, 4.1). Vor diesem Hintergrund soll nun das Datenmaterial auf das Vorkommen und die Qualität fachdidaktischen Wissens zur Planung von Unterricht zum experimentellen Handeln explorativ geprüft werden. Dabei werden zentrale empirische Befunde zum Forschungsanliegen 2a in den folgenden drei Unterkapiteln entlang der Forschungsfragen präsentiert.

3.1.1 Fachdidaktisches Wissen zum experimentellen Handeln in Unterrichtsplanungen

Die Forschungsfrage 1, ob angehende Lehrpersonen der Sekundarstufe I das experimentelle Handeln in ihren Unterrichtsplanungen im Sinne der Erkenntnisgewinnung umsetzen, lässt sich für die Gesamtstichprobe ($N = 119$) mit Hilfe der erreichten Summenscores, die auf der Grundlage der Ratings ermittelt worden sind, beantworten. Der Nachweis für die dafür notwendige Bedingung, dass die Skala eindimensional sein sollte (Döring & Bortz, 2016, S. 271), wurde in Teilstudie I bereits erbracht (Kap. I, 4.1.2). Die errechneten Summenscores sind in Tab. 15 dargestellt. Sie liegen aufgrund der verwendeten Rating-Skala von 1 bis 4 Punkten zwischen 17 (kein bzw. inadäquates experimentelles Handeln) und 68 Punkten (Höchstform experimentellen Handelns) und sind zur besseren Übersicht mittels Zehnerskala gruppiert worden.

Aus der Tabelle 15 kann entnommen werden, dass 70 der 119 untersuchten Unterrichtsplanungen, Arbeitsmaterialien sowie videografierten Planungsgespräche einen Summenscore grösser 17 Punkte verzeichnen. Das bedeutet, dass in insgesamt 59% aller Unterrichtsplanungen zumindest Ansätze von fachdidaktischem Wissen zur Planung von Unterricht zum experimentellen Handeln beobachtet werden konnten. Demgegenüber waren in 41% keinerlei Hinweise dazu zu finden. Eine genauere Betrachtung der Summenscores zeigt weiter, dass 10% aller Unterrichtsplanungen einen geringen Summenscore von 18-20 Punkten und nur 2% aller Unterrichtsplanungen einen Summenscore > 50 Punkte aufweisen. Die maximal erreichte Punktzahl beläuft sich bei der untersuchten Stichprobe ($N = 119$) auf 54 Punkte und liegt damit 14 Punkte unter der maximal erreichbaren Punktzahl. Eine genauere Betrachtung der Summenscores zeigt weiter, dass 10% aller Unterrichtsplanungen einen geringen Summenscore von 18-20 Punkten und nur 2% aller Unterrichtsplanungen einen Summenscore > 50 Punkte aufweisen.

Tabelle 25: Erreichte Summenscores zum experimentellen Handeln

Summenscore	17*	18 - 20	21 - 30	31 - 40	41 - 50	51 - 60
abs. Häufigkeit ($N = 119$)	49	12	28	14	13	3
				70		
rel. Häufigkeit ($N = 119$)	41%	10%	24%	12%	11%	2%
				59%		
Durchschnittlicher Summenscore ($N = 119$)						
	25.0 ($SD=10.3$)					
Durchschnittlicher Summenscore ($n = 70$)						
	30.6 ($SD=10.2$)					

* Ein Summenscore von 17 Punkten bedeutet, dass keine oder qualitativ schlechte Hinweise auf fachdidaktisches Wissen zur Planung von Unterricht zum experimentellen Handeln beobachtet worden sind.

Die maximal erreichte Punktzahl beläuft sich bei der untersuchten Stichprobe ($N = 119$) auf 54 Punkte und liegt damit 14 Punkte unter der maximal erreichbaren Punktzahl. Es fällt ausserdem auf, dass 24% aller Unterrichtsplanungen einen Summenscore zwischen 21 und 30 Punkten erzielen. Dies entspricht einem Bereich, in dem die mittleren Summenscores der Gesamt- ($N = 119$) und der Teilstichprobe ($n = 70$) liegen. Die durchschnittliche Punktzahl in der Gesamtstichprobe beträgt 25.0 Punkte ($SD = 10.3$), in der Teilstichprobe 30.6 Punkte ($SD = 10.2$). Letzteres deckt sich dabei annähernd mit einer mittleren Qualität fachdidaktischen Wissens zur Planung von Unterricht zum experimentellen Handeln.

Aufgrund dieser Ausführungen lässt sich zusammenfassend festhalten, dass in ca. 1/3 der 119 untersuchten Unterrichtsplanungen, -materialien sowie videografierten Planungsgespräche kein fachdidaktisches Wissen zur Planung von Unterricht zum experimentellen Handeln beobachtet werden konnte. In 2/3 der Stichprobe sind demgegenüber zumindest Hinweise zur Planung experimentellen Handelns festgestellt worden. Auf der Basis der Ratings ist damit eine Berechnung einer Kompetenzausprägung für die Planung von Unterricht zum experimentellen Handeln für Studierende möglich.

Die z-standardisierten Summenscores sind zur Veranschaulichung als Histogramme für die Gesamtstichprobe ($N = 119$, Abb. 27) und für die Teilstichprobe ($n = 70$, Abb. 28) dargestellt. Es ist anzumerken, dass für die absolute Häufigkeit jeweils eine passende Skala gewählt worden ist.

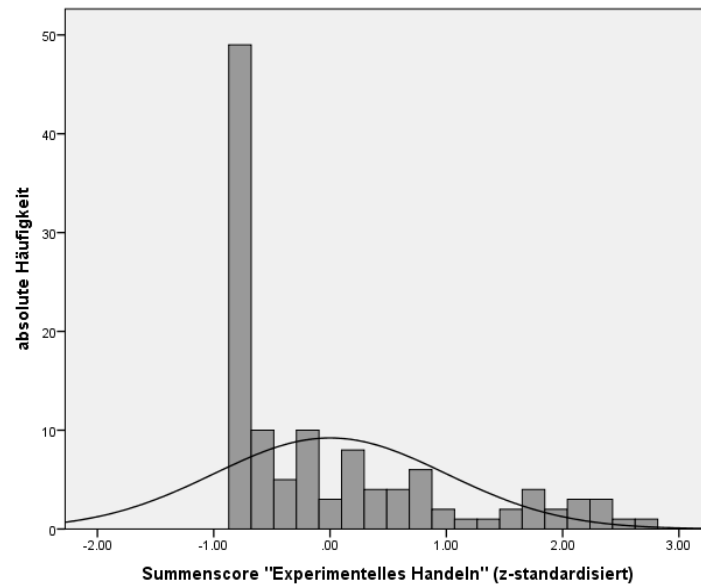


Abbildung 27: Kompetenzausprägungen zur Planung von Unterricht zum experimentellen Handeln, Gesamtstichprobe $N = 119$

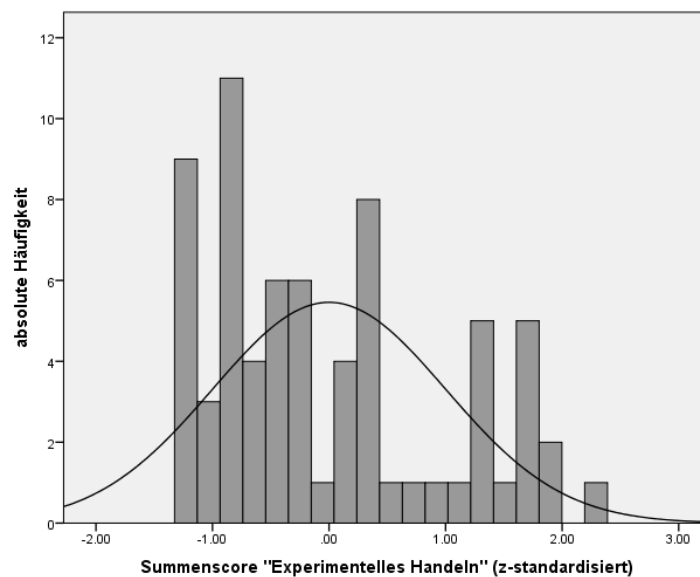


Abbildung 28: Kompetenzausprägungen zur Planung von Unterricht zum experimentellen Handeln, Teilstichprobe $n = 70$

Die in Abbildung 27 dargestellte Häufigkeitsverteilung der Kompetenzausprägungen der Gesamtstichprobe ($N = 119$) zeigt den bisher beschriebenen Bodeneffekt sehr deutlich. Bei 49 Studierenden konnten keinerlei Hinweise zum experimentellen Handeln in ihren Unterrichtsplanungen, -materialien und videografierten Planungsgesprächen beobachtet werden. Bei 70 Studierenden sind die Kompetenzausprägungen jedoch vergleichsweise höher, wobei nach wie vor tiefere Kompetenzausprägungen recht häufig sind. Dies zeigt sich auch

bei der zweiten Darstellung zu den Kompetenzausprägungen der Teilstichprobe ($n = 70$), bei der die Skala entsprechend der Häufigkeiten angepasst worden ist. In dieser Grafik verzeichnen die Kompetenzausprägungen eine stärkere Streuung und ist kein Bodeneffekt erkennbar (Abb. 28). Ob sich die zentralen Tendenzen der Teilstichprobe ($n = 70$) und der Reststichprobe ($n = 49$), in der keinerlei Hinweise auf experimentelles Handeln zu beobachten waren, signifikant unterscheiden, ist zunächst aufgrund nicht vorhandener Normalverteilung ($W = .91, p < 0.001$) und Varianzhomogenität ($F(1, 117) = 111.96, p < .001, N = 119$) nichtparametrisch mit Hilfe des Mann-Whitney-Tests überprüft worden (cf. Kap. II, 2.3). Die ermittelten Ergebnisse weisen auf einen signifikanten Unterschied zwischen den beiden Stichproben hin ($z = -9.61, p < .001$). Die Effektstärke entspricht mit $r = .88$ einem starken Effekt. Ein vergleichsweise ähnlicher Befund hat sich auch bei der parametrisch erfolgten Berechnung gezeigt ($t(117) = 11.09, p < .001, r = .71$). Daraus lässt sich ableiten, dass die 70 Studierenden, die das experimentelle Handeln in ihren Planungen zumindest ansatzweise gezeigt haben, sich signifikant von der anderen Gruppe unterscheiden und damit nicht alle Studierenden den in der Vignette verlangten Auftrag erfüllt haben. Auf der Basis der Gesamtstichprobe ($N = 119$) muss vor diesem Hintergrund die Hypothese verworfen werden, dass angehende Lehrpersonen, welche die Impulsveranstaltung I zum experimentellen Handeln besucht haben und einen entsprechenden Unterricht am Thema «Visuelle Wahrnehmung» planen müssen, diesen grösstenteils entsprechend gestalten.

Folglich werden empirische Befunde zur Forschungsfrage 2, in welcher Qualität angehende Lehrpersonen auf der Basis des im Projekt verwendeten Planungsrasters einen Unterricht zum experimentellen Handeln im Sinne naturwissenschaftlichen Erkenntnisgewinns planen, im nächsten Unterkapitel auf der Basis der Teilstichprobe $n = 70$ dargelegt.

3.1.2 Qualität fachdidaktischen Wissens zum experimentellen Handeln in Unterrichtsplanungen

Die deskriptiven Ergebnisse zur Forschungsfrage 2 sind in Abbildung 29 für die Teilstichprobe ($n = 70$) dargestellt. Die Grafik zeigt, dass die Mittelwerte der 5 Kategorien, die durch insgesamt 17 Indikatoren beschrieben sind, im Ratingmanual zwischen 1.34 und 2.18 liegen und bei einer Ratingskala von 1 bis 4 nicht allzu hoch sind. Die Kategorien «Fachliche Klärung (FK)» und «Unterrichtsziele (ZI)» weisen dabei die höchsten Mittelwerte, die Kategorien «Schülervorstellungen (SV)» und «Reflexivität (REF)» im Vergleich dazu die tiefsten Mittelwerte auf. Der Mittelwert der Kategorie «Didaktische Strukturierung (DS)» liegt dazwischen.

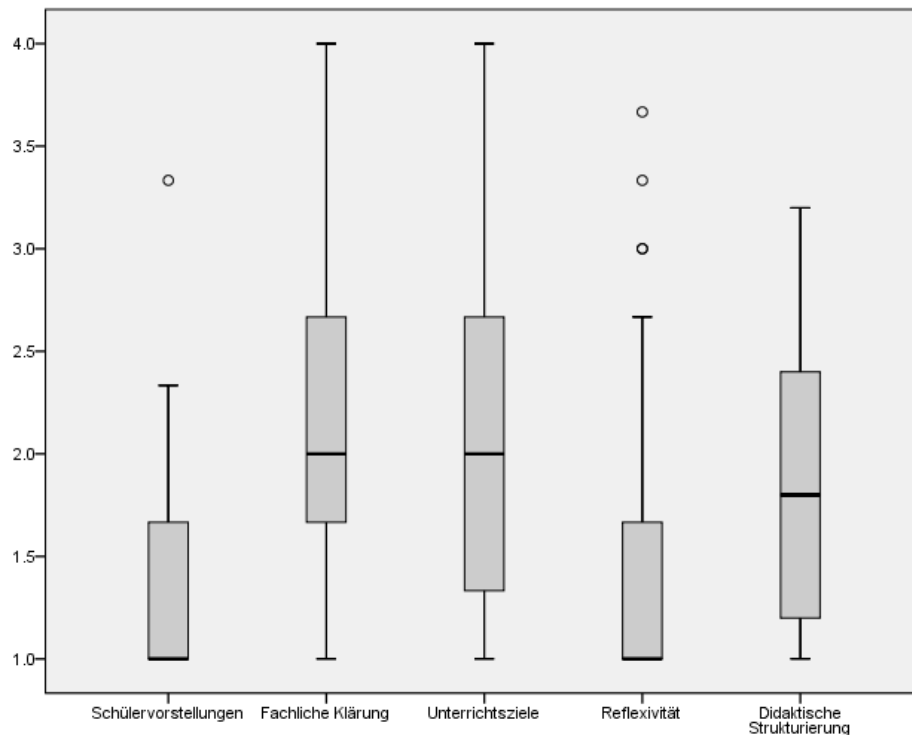


Abbildung 29: Mittelwertvergleiche der 5 Kategorien fachdidaktischen Wissens zur Planung von Unterricht zum experimentellen Handeln ($n = 70$)

Um genauere Aussagen zur Qualität, der im Ratingmanual verwendeten, 5 Kategorien machen zu können, sind zuerst die zentralen Tendenzen dieser Kategorien ermittelt worden (cf. Kap. I, 3.2.4). Die Daten sind, wie aus der Tabelle 16 entnommen werden kann, bei allen Kategorien nicht normalverteilt. Die Mittelwertvergleiche sind daher nichtparametrisch mit Hilfe des WILCOXON-Tests für abhängige Stichproben erfolgt. Für die parametrisch errechneten Ergebnisse sei auf den Anhang verwiesen (cf. Anhang B-1, Tab. B1.1)

Tabelle 16: Deskriptive Statistik der 5 Kategorien fachdidaktischen Wissens ($n = 70$)

Kategorien	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>	<i>Shapiro-Wilk-Test</i>
Schülervorstellungen (SV)	1.34	0.51	1.00	3.33	$W = 0.71^{***}, p < .001$
Fachliche Klärung (FK)	2.19	0.79	1.00	4.00	$W = 0.95^{**}, p < .01$
Unterrichtsziele (ZI)	2.08	0.86	1.00	4.00	$W = 0.92^{***}, p < .001$
Reflexivität (REF)	1.40	0.66	1.00	3.67	$W = 0.68^{***}, p < .001$
Didaktische Strukturierung (DS)	1.91	0.65	1.00	3.20	$W = 0.93^{**}, p < .01$

* $p < .05$, ** $p < .01$, *** $p < .001$

Die Ergebnisse der getesteten Mittelwertvergleiche, die in Tabelle 17 dargestellt sind, bestätigen die Vermutungen, die sich in Abbildung 29 andeuten. Diejenigen Kategorien, deren Mittelwerte am nächsten beieinander liegen, unterscheiden sich nicht signifikant, wohingegen alle anderen Mittelwertvergleiche statistisch hoch signifikant sind. Die Effektstärken sind dabei mittel bis stark. Dieses Bild bestätigt sich durch die parametrisch berechneten Mittelwertvergleiche (Anhang B-1, Tab. B1.1).

Tabelle 17: Mittelwertvergleich der 5 Kategorien ($n = 70$)

Kategorien	<i>z</i>	<i>p</i>	<i>r</i>
Schülervorstellungen (SV) – Fachliche Klärung (FK)	-6.78	< .001	.81
Schülervorstellungen (SV) – Unterrichtsziele (ZI)	-6.39	< .001	.76
Schülervorstellungen (SV) – Didaktische Strukturierung (DS)	-6.42	< .001	.77
Schülervorstellungen (SV) – Reflexivität (REF)	-.08	.44	.09
Fachliche Klärung (FK) – Unterrichtsziele (ZI)	-1.48	.14	.18
Fachliche Klärung (FK) – Didaktische Strukturierung (DS)	-3.80	< .001	.45
Fachliche Klärung (FK) – Reflexivität (REF)	-6.54	< .001	.78
Unterrichtsziele (ZI) – Didaktische Strukturierung (DS)	-2.82	.005	.34
Unterrichtsziele (ZI) – Reflexivität (REF)	-6.24	< .001	.75
Didaktische Strukturierung (DS) – Reflexivität (REF)	-5.90	< .001	.71

Aus diesen empirischen Befunden kann abgeleitet werden, dass in Bezug auf das fachdidaktische Wissen zu den Kategorien «Fachliche Klärung», «Unterrichtsziele» und «Didaktische Strukturierung» in den Unterrichtsplanungen, -materialien und videografierten Planungsgesprächen Hinweise in besserer Qualität vorliegen, während zwischen den Kategorien «Fachliche Klärung» und «Unterrichtsziele» kein signifikanter Unterschied besteht. Demgegenüber sind zu den beiden Kategorien «Schülervorstellungen» und «Reflexivität» Hinweise in geringerer Qualität festgestellt worden. Die Mittelwerte dieser beiden Kategorien zeigen ebenfalls keinen signifikanten Unterschied.

In der anschliessend erfolgten Analyse der Interkorrelationen der 5 Kategorien erweisen sich alle Korrelationen auch nach erfolgter Bonferroni-Korrektur als signifikant. Die Korrelationen der Kategorien «Schülervorstellungen» und «Reflexivität» sind jedoch etwas tiefer als die der übrigen Kategorien (Tab. 18). Alle Korrelationen liegen gemäss Weiber und Mühlhaus (2014) im mittleren bis hohen Bereich. Die parametrisch ermittelten Korrelationswerte zeigen ein ähnliches Bild (Anhang B-2, Tab. B1.2). Daraus kann abgeleitet

werden, dass das deklarative fachdidaktische Wissen bei den 5 Kategorien jeweils in einer ähnlichen Qualität beobachtet werden konnte und sich dies auf die Kompetenzausprägung der latenten Variablen ausgewirkt hat.

Tabelle 18: Interkorrelationen der 5 Kategorien ($n = 70$)

Kategorien	SV	FK	ZI	REF	DS
Schülervorstellungen (SV)	–	.65**	.64**	.65**	.60**
Fachliche Klärung (FK)		–	.76**	.70**	.74**
Unterrichtsziele (ZI)			–	.69**	.82**
Reflexivität (REF)				–	.65**
Didaktische Strukturierung (DS)					–

** $p < .01$

Ob dieser Befund Auswirkungen hat auf die beiden Dimensionen des Modells der «Didaktischen Rekonstruktion», «Analyse» und «Konstruktion», die bei der Validierung des Instruments ermittelt werden konnten, ist in einem weiteren Schritt geprüft worden. Die zentralen Tendenzen der beiden Dimensionen sind in Tabelle 19 dargestellt.

Tabelle 19: Deskriptive Statistik der zwei Dimensionen «Analyse» und «Konstruktion» ($n = 70$)

Kategorien	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>	<i>Shapiro-Wilk-Test</i>
Analyse (ANA)	1.85	0.75	1.00	3.56	$W = 0.88^{***}, p < .001$
Konstruktion (KON)	1.98	0.70	1.00	3.60	$W = 0.94^{**}, p < .01$

* $p < .05$, ** $p < .01$, *** $p < .001$

Auch die Daten beider Dimensionen sind nicht normalverteilt. Der Mittelwertvergleich zwischen den beiden Dimensionen ist daher primär nichtparametrisch mit Hilfe des WILCOXON-Tests für abhängige Stichproben erfolgt. Dieser zeigt, dass beide Dimensionen statistisch signifikant verschieden sind ($z = -2.69, p = .007$). Die Effektstärke nach Cohen (1992) beträgt $r = .32$ und entspricht damit einem mittleren Effekt. Ein entsprechendes Ergebnis hat sich auch mit Hilfe des parametrischen t-Tests für abhängige Stichproben ergeben ($t(69) = -2.152, p = .035, r = .25$). Das bedeutet, dass die Qualität des beobachteten fachdidaktischen Wissens der Dimension «Konstruktion» im Vergleich zur Dimension «Analyse» in den Planungen in besserer Qualität umgesetzt worden ist. Dies ist nicht weiter

erstaunlich, da 4 der 9 Indikatoren der Dimension «Analyse» den beiden Kategorien «Schülervorstellungen» und «Reflexivität» entstammen. Die in der Folge nichtparametrisch geprüfte Interkorrelation zeigt weiter, dass die beiden Dimensionen signifikant miteinander korrelieren ($r_s = .77, p < .01$). Die parametrisch ermittelte Korrelation liegt bei $r = .74, p < .01$). Die Korrelationswerte liegen im hohen Bereich (Weiber & Mühlhaus, 2014, S. 15). Dieses Ergebnis kann dahingehend gedeutet werden, dass die beiden Dimensionen in den Planungen jeweils in ähnlicher Qualität zu beobachten waren und sich dies entsprechend auch auf die Kompetenzausprägung ausgewirkt hat.

Vergleicht man die Rubriken im Planungsraster mit den hier erzielten Erkenntnissen, so lässt sich festhalten, dass die drei Kategorien, die in den Unterrichtsplanungen qualitativ besser umgesetzt worden sind, im Planungsraster (cf. Anhang A-1, Abb. A1.2) explizit aufgeführt waren. Vor diesem Hintergrund kann vermutet werden, dass die drei Kategorien, die im Planungsraster vorgegeben waren, sich auf die Planungsüberlegungen ausgewirkt haben. Hinweise zu Planungsüberlegungen zu Schülervorstellungen, aber auch zur Reflexivität sind weniger und in geringerer Qualität in den Planungen beobachtbar. Entsprechend widerspiegelt sich dieser Befund auch in der Qualität der Dimension «Analyse». Insofern kann die Hypothese gestützt werden, dass die Qualität der Planungsüberlegungen durch das Planungsinstrument beeinflusst wird, wenn angehende Lehrpersonen auf der Basis des vorgegebenen Planungsrasters einen Unterricht zum experimentellen Handeln planen.

Ausgehend von diesen Erkenntnissen widmet sich das nächste Unterkapitel der Forschungsfrage 3, bei der die vergleichende Untersuchung der Interventions- und Kontrollgruppe im Zentrum steht. Es werden hierzu zentrale empirische Befunde vorgestellt.

3.1.3 Qualität in Unterrichtsplanungen – ein Vergleich von Interventions- und Kontrollgruppe

Empirische Befunde zur Forschungsfrage 3, ob sich die im Projekt KUBeX untersuchten Interventions- und Kontrollgruppen hinsichtlich der Qualität ihrer Unterrichtsplanungen zum experimentellen Handeln unterscheiden, sind in Tabelle 21 für die Gesamtstichprobe ($N = 119$) dargestellt. Das Augenmerk der Analyse lag auf der Ermittlung möglicher Unterschiede hinsichtlich der latenten Variablen «Experimentelles Handeln», aber auch in Bezug auf die beiden Dimensionen «Analyse» (ANA) und «Konstruktion» (KON). Aufgrund fehlender Normalverteilung der Daten (cf. Tab. 20) ist die Analyse nichtparametrisch mit Hilfe des Mann-Whitney-U-Tests erfolgt.

Tabelle 20: Deskriptive Statistik der zwei Dimensionen fachdidaktischen Wissens und der latenten Variablen (expH) für die Interventionsgruppe (IG, $n = 65$) und Kontrollgruppe (KG, $n = 54$)

Dimensionen	<i>M</i>	<i>SD</i>	Shapiro-Wilk-Test	<i>M</i>	<i>SD</i>	Shapiro-Wilk-Test
	IG ($n = 65$)			KG ($n = 54$)		
Analyse (ANA)	1.38	0.67	$W = 0.63^{***}, p < .001$	1.65	0.74	$W = 0.83^{***}, p < .001$
Konstruktion (KON)	1.48	0.70	$W = 0.73^{***}, p < .001$	1.70	0.73	$W = 0.85^{***}, p < .001$
Exp. Handeln (expH)	1.41	0.66	$W = 0.69^{***}, p < .001$	1.66	0.71	$W = 0.86^{***}, p < .001$

* $p < .05$, ** $p < .01$, *** $p < .001$

Die Ergebnisse in Tabelle 21 weisen darauf hin, dass sich die Interventions- und Kontrollgruppe in Bezug auf die Qualität fachdidaktischen Wissens zur Planung von Unterricht zum experimentellen Handeln signifikant unterscheiden, wobei in den Arbeiten der Kontrollgruppe eine etwas bessere Qualität feststellbar war. Ein signifikanter Unterschied ist auch für die Dimension der «Analyse» zu verzeichnen, wohingegen kein signifikanter Unterschied in Bezug auf die Dimension «Konstruktion» ermittelt werden konnte.

Tabelle 21: Mittelwertvergleiche der Dimensionen (ANA & KON) und der latenten Variablen (expH) für die Interventions- (IG) und Kontrollgruppe (KG)

Dimensionen	<i>z</i>	<i>p</i>	<i>r</i>
Analyse (IG) – Analyse (KG)	-2.58	.010*	.24
Konstruktion (IG) – Konstruktion (KG)	-1.89	.059	.17
Exp. Handeln (IG) – exp. Handeln (KG)	-2.41	.016*	.22

* $p < .05$, ** $p < .01$, *** $p < .001$

Alle Effektstärken sind gemäss Cohen (1992) allerdings eher schwach bis mittel. Gestützt auf diese Ergebnisse kann vermutet werden, dass die signifikanten Unterschiede zwischen beiden Untersuchungsgruppen vorwiegend auf Indikatoren aus der Dimension der «Analyse» zurückzuführen sind. Die vor diesem Hintergrund berechneten Mittelwertvergleiche der in der Dimension «Analyse» enthaltenen Indikatoren bzw. Kategorien sind in Bezug auf Unterschiede zwischen Interventions- und Kontrollgruppe daher weiter geprüft worden. Aufgrund der berechneten zentralen Tendenzen der Daten, erfolgte dies erneut nichtparametrisch (Tab. 22).

Tabelle 22: Deskriptive Statistik der Kategorien der Dimension «Analyse» für die Interventionsgruppe (IG, $n = 65$) und Kontrollgruppe (KG, $n = 54$)

Dimensionen	<i>M</i>	<i>SD</i>	Shapiro-Wilk-Test	<i>M</i>	<i>SD</i>	Shapiro-Wilk-Test
	IG ($n = 65$)			KG ($n = 54$)		
ANA-SV	1.20	0.54	$W = 0.42^{***}, p < .001$	1.37	0.63	$W = 0.64^{***}, p < .001$
ANA-FK	1.50	0.83	$W = 0.66^{***}, p < .001$	1.84	0.84	$W = 0.58^{***}, p < .001$
ANA-ZI	1.55	0.90	$W = 0.66^{**}, p < .001$	1.96	0.86	$W = 0.86^{***}, p < .001$
ANA-REF	1.22	0.57	$W = 0.43^{***}, p < .001$	1.40	0.58	$W = 0.86^{***}, p < .001$
ANA-DS	1.48	0.70	$W = 0.73^{***}, p < .001$	1.69	0.86	$W = 0.86^{***}, p < .001$

* $p < .05$, ** $p < .01$, *** $p < .001$

Aus den Befunden in Tabelle 23 lässt sich entnehmen, dass die Kontrollgruppe in den Kategorien «Unterrichtsziele», «Fachliche Klärung» und «Schülervorstellungen» eine signifikant bessere Qualität zeigt, in den anderen beiden Kategorien dieser Dimension jedoch nicht. Die Effektstärken bewegen sich auch hier im schwachen bis mittleren Bereich. Analysen zur Dimension «Konstruktion» hinsichtlich der darin enthaltenen Indikatoren bzw. Kategorien und Indikatoren haben keine signifikanten Unterschiede gezeigt (cf. Anhang B-1, Tab. B1.3 & B1.4)

Tabelle 23: Mittelwerte der Kategorien der Dimension «Analyse» im Vergleich ($n = 70$)

Dimensionen	<i>z</i>	<i>p</i>	<i>r</i>
Analyse (ANA) – Schülervorstellungen (SV)	-2.17	.030*	.019
Analyse (ANA) – Fachliche Klärung (FK)	-2.41	.016*	.022
Analyse (ANA) – Unterrichtsziele (ZI)	-2.70	.007**	.025
Analyse (ANA) – Reflexivität (REF)	-1.15	.251	.011
Analyse (ANA) – Didaktische Strukturierung (DS)	-1.89	.059	.017

* $p < .05$, ** $p < .01$, *** $p < .001$

Ausserdem hat sich auch die Vermutung, dass sich die Interventions- und Kontrollgruppe in Bezug auf die Ergebnisse der Professionswissenstests (Fachwissen, CK (t1) sowie Fachdidaktisches Wissen (PCK (t2))) unterscheiden, nicht bestätigt (cf. Anhang B-1, Tab. B1.5).

Auf der Basis der hier analysierten Daten kann die Hypothese nicht gestützt werden, dass die Interventionsgruppe aufgrund der in einer Intervention erfolgten und vertieften Auseinandersetzung mit den Kernaspekten einen qualitativ besseren Unterricht zum experimentellen Handeln als die Kontrollgruppe plant.

Ausgehend von den bisherigen Analysen zum Vorkommen und zur Qualität der Planungen zum experimentellen Handeln widmet sich das nächste Kapitel (Kap. II, 3.2) dem Forschungsanliegen 2b und damit möglichen Zusammenhängen zwischen der Planungsqualität und den Ergebnissen der Professionswissenstests.

3.2 Zusammenhänge zwischen der Planungsqualität und den Ergebnissen der Professionswissenstests

Zur Ermittlung möglicher Wirkungszusammenhänge zwischen dem in den Planungen erfassten, handlungsnäheren fachdidaktischen Wissen (Performanz) und dem im Rahmen von Professionswissenstests erfassten Testwissen (Disposition) sind bivariate Korrelationen berechnet worden. Die Ergebnisse konnten auch bei den Daten der Tests ohne fehlende Werte ermittelt werden und sind ebenfalls nichtparametrisch (Spearman-Korrelation) wie auch parametrisch (Pearson-Korrelation) berechnet worden. Die nichtparametrisch berechneten Ergebnisse werden in den nächsten beiden Unterkapiteln vorgestellt, parametrisch ermittelte Korrelationen sind im Anhang aufgeführt. In einem ersten Schritt werden die Ergebnisse der Zusammenhänge zwischen dem im Test erfassten fachdidaktischen Wissen und der Qualität der Planungen präsentiert (Kap. II, 3.2.1). Die Ergebnisse zu einem möglichen Zusammenhang zwischen dem getesteten Fachwissen und der Planungsqualität werden in einem zweiten Schritt aufgezeigt (Kap. II, 3.2.2). Alle Berechnungen sind mit den Daten der Gesamtstichprobe ($N = 119$) erfolgt.

3.2.1 Einfluss des deklarativen fachdidaktischen Wissens auf die Planungsqualität

Ergebnisse zur Forschungsfrage 4, ob sich Zusammenhänge zwischen dem deklarativen fachdidaktischen Wissen und der Qualität der Unterrichtsplanungen in Bezug auf das experimentelle Handeln im Biologieunterricht feststellen lassen, sind in Tabelle 24 für die Gesamtstichprobe ($N = 119$) dargestellt.

Tabelle 24: Korrelationen zwischen dem getesteten fachdidaktischen Wissen (FDW) und dem in Planungsmaterialien erfassten fachdidaktischen Wissen (expH) ($N = 119$)

Kategorien	<i>M</i>	<i>SD</i>	FDW	expH	ANA	KON
Fachdidaktisches Wissen (FDW)	0.59	0.87	–	.41**	.39**	.43**

** $p < .01$

Die nichtparametrisch ermittelten Ergebnisse lassen erkennen, dass das im Test erhobene fachdidaktische Wissen zur Planung von Unterricht zum experimentellen Handeln signifikant mit der Qualität fachdidaktischen Wissens korreliert, das in den Unterrichtsplanungen, -materialien und videografierten Planungsgesprächen erfasst werden konnte, ($r_s = .41$, $p < 0.001$, $N = 119$). Weiter ist auch ein signifikanter Zusammenhang zwischen dem fachdidaktischen Wissen und der Dimension «Analyse» ($r_s = .39$, $p < 0.001$, $N = 119$) bzw. der Dimension «Konstruktion» ($r_s = .43$, $p < 0.001$, $N = 119$) zu verzeichnen. Die Effektstärken der Korrelationen liegen damit alle im mittleren bis starken Bereich (Cohen, 1992; cf. Field, 2013). Das bedeutet, dass diejenigen Studierenden, die im fachdidaktischen Professionswissenstest besser abgeschnitten, entsprechend auch qualitativ besseren Unterricht zum Experimentieren im Sinne der Erkenntnisgewinnung geplant haben. Darüber hinaus weisen diese Ergebnisse darauf hin, dass die entsprechenden Kompetenzausprägungen in Bezug auf die beiden Planungsdimensionen «Analyse» und «Konstruktion» mit den Testresultaten korrelieren. Ein entsprechender Befund zeigt sich auch bei den parametrisch ermittelten Korrelationen (cf. Anhang B-2, Tab. B2.1). Die Hypothese, dass angehende Lehrpersonen, die im Test ein gutes fachdidaktisches Wissen zeigen, einen qualitativ besseren Unterricht zum experimentellen Handeln als Fachinhalt planen, kann auf der Basis der erhobenen Daten und auf der Grundlage dieser Befunde bestätigt werden.

Vor diesem Hintergrund soll im nächsten Unterkapitel die Forschungsfrage 5, nämlich wie sich die Ergebnisse aus dem Fachwissenstest zur visuellen Wahrnehmung auf die Planungsqualität auswirken, mit empirischen Belegen unterfüttert werden.

3.2.2 Einfluss des deklarativen Fachwissens auf die Planungsqualität

Empirische Befunde zur Forschungsfrage 5, ob sich Zusammenhänge zwischen dem deklarativen Fachwissen und der Qualität der Unterrichtsplanungen in Bezug auf das experimentelle Handeln im Biologieunterricht feststellen lassen, sind ebenfalls für die Gesamtstichprobe ($N = 119$) in Tabelle 25 aufgeführt. Die nichtparametrisch ermittelten Ergebnisse zeigen, dass die Korrelation zwischen dem Fachwissen und der erfassten Qualität in den Planungsmaterialien knapp signifikant ist ($r_s = .19$, $p = 0.036$, $N = 119$). Dies gilt auch

für die Korrelation zwischen dem Fachwissen und den beiden Dimensionen «Analyse» ($r_s = .19, p = 0.038, N = 119$) und «Konstruktion» ($r_s = .18, p = 0.046, N = 119$). Die Effektstärken der Korrelationen liegen jedoch im schwachen bis mittleren Bereich (Cohen, 1992; cf. Field, 2013).

Tabelle 25: Korrelationen zwischen dem getesteten Fachwissen (FW) und dem in Planungsmaterialien erfassten fachdidaktischen Wissen (expH) sowie den beiden Dimensionen «Analyse» und «Konstruktion» (nichtparametrisch und parametrisch; $N = 119$)

Kategorien	<i>M</i>	<i>SD</i>	FW	expH	ANA	KON
Fachwissen (FW) nichtparametrisch	-.031	0.87	–	.19*	.19*	.18*
Fachwissen (FW) parametrisch	-.031	0.87	–	.12	.10	.15

* $p < .05$ (2-seitig)

Diese Ergebnisse sind jedoch nicht kongruent mit den parametrisch errechneten Korrelationen (Tab. 25). Die Korrelation zwischen dem getesteten Fachwissen und der Qualität des fachdidaktischen Wissens in den Planungsunterlagen ist nicht signifikant ($r = .12, p = 0.178, N = 119$), genauso wenig wie die Korrelationen des Fachwissens mit der Dimension «Analyse» ($r = .10, p = 0.273, N = 119$) bzw. der Dimension «Konstruktion» ($r = .15, p = 0.099, N = 119$). Es hat sich darüber hinaus gezeigt, dass keine signifikante Korrelation zwischen dem getesteten fachdidaktischen Wissen und dem erhobenen Fachwissen besteht ($r_s = 0.025, p = 0.79, N = 119$). Dies deutet darauf hin, dass die beiden Wissensdimensionen keinen direkten Bezug zueinander aufweisen. Vor dem Hintergrund, dass die nichtparametrische Spearman-Korrelation auf Rangordnungen basiert und Effekte extremer Werte wie auch Verletzungen hinsichtlich Varianzhomogenität und Normalverteilung minimiert (Field, 2013), ist dieses Verfahren für die Daten dieser Studie aufgrund fehlender Normalverteilung grundsätzlich vorzuziehen. Auch die Tatsache, dass die mittels Bootstrap ermittelten Konfidenzintervalle die Nullgrenze nicht überschreiten, stützt gemäß Field (2013) die signifikanten Effekte. Diese sind aber nur schwach bis mittel einzustufen, insbesondere auch bei den beiden Dimensionen «Analyse» und «Konstruktion». Die Hypothese, dass das im Projekt erhobene Fachwissen einen Einfluss auf die Qualität der Unterrichtsplanungen hat, kann deshalb nur mit grossem Vorbehalt gestützt werden.

3.3 Zusammenfassung der Ergebnisse der Teilstudie II

Ziel der in Kapitel II, 3 erfolgten deskriptiven und inferenzstatistischen Analyse war die explorative Prüfung von Unterrichtsplanungen, -materialien und videografierten Planungsgesprächen angehender Lehrpersonen in Bezug auf den darin angelegten Unterricht zum

experimentellen Handeln. Ausserdem wurden mögliche Zusammenhänge zwischen dem im Projekt KUBeX getesteten fachdidaktischen Wissen und Fachwissen sowie dem in den Planungen erfassten, handlungsnäheren fachdidaktischen Wissen geprüft. Die Ergebnisse zur Forschungsfrage 1 zeigen, dass nur bei 70 aller 119 Planungsmaterialien Hinweise zum Experimentieren im Sinne der Erkenntnisgewinnung zumindest ansatzweise feststellbar sind. Diese Teilstichprobe ($n = 70$) unterscheidet sich ausserdem signifikant von der Reststichprobe ($n = 49$) und weist unterschiedliche Kompetenzausprägungen der themenspezifischen Planungskompetenz auf. Tiefere Kompetenzausprägungen sind auch in der Teilstichprobe ($n = 70$) nach wie vor recht häufig. Befunde zur Forschungsfrage 2, die an der Teilstichprobe ($n = 70$) untersucht worden ist, deuten weiter darauf hin, dass das fachdidaktische Wissen zu den Kategorien «Fachliche Klärung», «Unterrichtsziele» und «Didaktische Strukturierung» in den Planungsmaterialien in besserer Qualität vorliegt und Hinweise zu den Kategorien «Schülervorstellungen» und «Reflexivität» in geringerer Qualität festzustellen sind. Vergleicht man die Rubriken im Planungsraster mit den hier erzielten Erkenntnissen, so lässt sich festhalten, dass die drei Kategorien, die in den Unterrichtsplanungen qualitativ besser umgesetzt worden sind, im Planungsraster explizit aufgeführt waren. Gestützt auf die ermittelten Interkorrelationen der fünf Kategorien sowie der Korrelationen der Kategorien mit der latenten Variablen «Experimentelles Handeln» (expH), die alle signifikant sind, kann abgeleitet werden, dass die Qualität deklarativen fachdidaktischen Wissens bei allen Kategorien jeweils ähnlich ist und sich dies daher auch auf die Kompetenzausprägung der latenten Variablen ausgewirkt hat. Entsprechend sind auch Auswirkungen auf die beiden Dimensionen des Modells der «Didaktischen Rekonstruktion», «Analyse» und «Konstruktion», zu verzeichnen. Die Qualität des beobachteten fachdidaktischen Wissens der Dimension «Konstruktion» ist hier im Vergleich zur Dimension «Analyse» in den Planungsmaterialien höher. Dieses Ergebnis wird auch durch die signifikanten Interkorrelationen zwischen den beiden Dimensionen und der latenten Variablen gestützt, was sich entsprechend auf die Kompetenzausprägungen auswirkt. Analysen zur Forschungsfrage 3 zeigen ausserdem, dass in den Planungsmaterialien der Kontrollgruppe im Vergleich zur Interventionsgruppe eine signifikant bessere Qualität fachdidaktischen Wissens zur Planung von Unterricht zum experimentellen Handeln festgestellt werden konnte. Die Effektstärken bewegen sich allerdings im schwachen bis mittleren Bereich. Ein entsprechendes Bild zeigt sich auch im Zusammenhang mit der in den Planungsmaterialien beobachteten Qualität der beiden Dimensionen «Analyse» und «Konstruktion». Die signifikanten Unterschiede lassen sich hier vorwiegend auf Indikatoren der Dimension «Analyse» zurückführen. Dabei weisen Indikatoren der Kategorien «Unterrichtsziele», «Fachliche Klärung» und «Schülervorstellungen» in den Planungsmaterialien der Kontrollgruppe eine signifikant bessere Qualität auf. Dieser Unterschied besteht allerdings nicht in den

anderen beiden Kategorien «Schülervorstellungen» und «Reflexivität». Befunde zur Forschungsfrage 4 zeigen, dass das im Test erhobene deklarative fachdidaktische Wissen zur Planung von Unterricht zum experimentellen Handeln stark mit der Qualität des deklarativen fachdidaktischen Wissens in den Unterrichtsplanungen, -materialien und videograferten Planungsgesprächen korreliert. Das bedeutet, dass diejenigen Studierenden, die im fachdidaktischen Professionswissenstest besser abgeschnitten haben, entsprechend auch qualitativ besseren Unterricht zum Experimentieren im Sinne der Erkenntnisgewinnung geplant haben. Dieses Ergebnis widerspiegelt sich auch in den untersuchten Korrelationen des getesteten deklarativen fachdidaktischen Wissens mit dem erfassten deklarativen fachdidaktischen Wissen beider Dimensionen. Ein signifikanter Zusammenhang zwischen dem getesteten deklarativen Fachwissen und der latenten Variablen «Experimentelles Handeln» bzw. den beiden Dimensionen «Analyse» und «Konstruktion» konnte im Rahmen der fünften und letzten Forschungsfrage 5 ermittelt werden. Dieser eher erstaunliche Befund, wie auch alle anderen hier beschriebenen Ergebnisse werden im nächsten Kapitel vor dem Hintergrund des aktuellen Forschungsstandes diskutiert und eingeordnet.

4 Diskussion

Mit den neuen Bildungsstandards (EDK, 2011; KMK, 2005) sind (angehende) Lehrpersonen verpflichtet, Kompetenzen im Bereich des Experimentierens im Sinne der Erkenntnisgewinnung zu fördern. Damit verbunden ist auch die Erarbeitung eines Verständnisses der durch Konventionen und ethische Grundsätze gekennzeichneten Prozesse, mit denen naturwissenschaftliche Erkenntnisse generiert werden (Roberts, 2007; Schwartz et al., 2004). Für die Gestaltung eines solch komplexen Unterrichts sollten (angehende) Lehrpersonen über ein fundiertes themenspezifisches Professionswissen und die damit verbundenen professionellen Handlungskompetenzen verfügen (Crawford, 2000, 2007), insbesondere auch im Biologieunterricht. Bisherige Studien zeigen, dass es Lehrpersonen dafür u.a. an fachlichem Wissen und Verständnis des Konzepts *Scientific Inquiry*, aber auch an fachdidaktischem Wissen mangelt (e.g. Gyllenpalm & Wickman, 2011b; Gyllenpalm et al., 2010). Das Planungshandeln bietet einen geeigneten Anlass, das fachdidaktische Wissen in einer handlungsnahen Situation anzuwenden (Smith & Banilower, 2015) und aufzubauen (Brown, Friedrichsen, & Abell, 2013). Allerdings scheint es, dass Planungsraster im Gegensatz zu Planungsmodellen die Planungsüberlegungen von Berufsanfängerinnen und Berufsanfängern stärker einschränken und zu weniger Planungsüberlegungen führen (Zierer, Werner & Wernke, 2015). Bisher gibt es nur wenige Studien mit einem naturwissenschaftsdidaktischen Fokus, die Aussagen zur Planungskompetenz und damit zur Qualität der Unterrichtsplanungen erlauben (Stender, 2014; Stender et al., 2017). Diese sind jedoch fachlich und nicht fachmethodisch ausgerichtet. Ob es (angehenden) Lehrpersonen gelingt, einen Unterricht zum Experimentieren im Sinne der Erkenntnisgewinnung zu planen, ist daher noch kaum erforscht (Schneider & Plasman, 2011). Auch ist wenig bekannt, welche Qualität unterrichtliche Planungen von Studierenden zum Experimentieren im Sinne der Erkenntnisgewinnung erreichen und durch welche Faktoren (Fachwissen und fachdidaktisches Wissen zum Unterrichten naturwissenschaftlicher Erkenntnisgewinnung) die Planungsqualität beeinflusst wird. Vor diesem Hintergrund sind die im Projekt KUBeX von angehenden Lehrpersonen erstellten Unterrichtsplanungen, -materialien und videografierten Planungsgespräche mit Hilfe des in Teilstudie I inhaltlich und faktoriell validierten Ratingmanuals untersucht worden. Ziel der Teilstudie II ist eine erste explorative Untersuchung der im Projekt erhobenen Planungsunterlagen sowie möglicher Zusammenhänge zu den Ergebnissen der Professionswissenstests. Die in Kapitel II, 3 dargelegten Ergebnisse werden ausgehend vom aktuellen Forschungsstand in den nächsten Abschnitten diskutiert. Dabei werden auch Limitationen der Teilstudie II aufgezeigt.

4.1 Befunde zum Vorkommen und zur Qualität experimentellen Handelns in den Planungsmaterialien (Forschungsanliegen 2a)

Befunde zum Forschungsanliegen 2a zeigen, dass Anstösse zu Lerngelegenheiten, die das Erlernen der Teilkompetenzen wissenschaftlichen Denkens im Unterricht fördern könnten, nur in 70 der 119 untersuchten Planungsmaterialien beobachtet werden konnten und dies trotz der besuchten Impulsveranstaltung I (2×90 Minuten) zum Experimentieren im Sinne der Erkenntnisgewinnung. Weiter hat die Analyse ergeben, dass bei der Teilstichprobe ($n = 70$), die das experimentelle Handeln in den Planungsmaterialien zumindest ansatzweise umgesetzt hat, tiefere Kompetenzausprägungen vergleichsweise häufiger waren. Diese beiden Ergebnisse deuten darauf hin, dass die Studierenden der untersuchten Stichprobe Schwierigkeiten bekundet haben, einen entsprechenden Unterricht zu gestalten. Die ermittelten Korrelationen zwischen dem getesteten fachdidaktischen Wissen und dem in den Planungsmaterialien erfassten fachdidaktischen Wissen stützen diese Interpretation. Entsprechend könnten, wie in der Literatur mehrfach erwähnt (vgl. Fazio et al., 2010; Roehrig & Luft, 2004; Yoon et al., 2012, Kap. 1.2) diesbezügliche fehlende Fähigkeiten und Fertigkeiten der Studierenden auf die Qualität der Planungsmaterialien abgefärbt haben. Die in Kapitel 5.3 der Teilstudie I angesprochene Konfundierung mit anderen Planungsaufträgen, die aufgrund des gewählten Settings nicht vermieden werden konnte, hat hier möglicherweise auch die Planungsqualität beeinflusst. Vor dem Hintergrund, dass Stender et al. (2015) bei angehenden Lehrpersonen einen direkten Zusammenhang zwischen der Motivation sowie den selbstregulativen Fähigkeiten und der Qualität der Skripts nachweisen konnten, wäre auch denkbar, dass die Motivation Einfluss auf die Qualität der gestalteten Planungen genommen hat. Denn die Studierenden mussten den im Projekt KUBeX eingesetzten Planungsauftrag losgelöst von anderen Planungsaufträgen umsetzen und konnten ausserdem diese Planung aus testökonomischen Gründen nicht einmal realisieren. Nicht vergessen darf man hier aber auch den Umstand, dass die Erfassung der Kompetenzausprägung auf einer einzigen Planung beruht. Wie Praetorius mit anderen Autoren hinsichtlich der Analyse von Unterricht aufzeigen konnte, bedarf es für eine zuverlässige Einschätzung der Unterrichtsqualität hinsichtlich der kognitiven Aktivierung (Kriterium der Tiefenstruktur) insgesamt 9 Lektionen (Praetorius, Pauli, Reusser, Rakoczy & Klieme, 2014, S. 8). Mit nur einer einzigen Lektion ist zwar bereits eine angemessene Interrater-Reliabilität erreicht, aber eine solche Einschätzung der Unterrichtsqualität sollte nicht als Prädiktor für die Lernleistung der Schülerinnen und Schüler herangezogen werden (Hugener, Pauli & Reusser, 2007). Es kann vermutet, dass Ähnliches für die Qualität von Unterrichtsplanungen gilt.

Genauere Analysen des erfassten fachdidaktischen Wissens in den Planungsunterlagen und videografierten Planungsgesprächen haben weiter gezeigt, dass das fachdidaktische Wissen aus den Kategorien «Fachliche Klärung», «Unterrichtsziele» und «Didaktische Strukturierung» im Vergleich zu den Kategorien «Schülervorstellungen» und «Reflexivität» in besserer Qualität vorlag. Entsprechend hat sich dies auch auf die Qualität der beiden Dimensionen der «Didaktischen Rekonstruktion» ausgewirkt, wobei das fachdidaktische Wissen zur Dimension «Konstruktion» qualitativ besser war. Ein Vergleich mit den im Planungsraster explizit aufgeführten Rubriken lässt hierzu vermuten, dass diese die Planungsüberlegungen der Studierenden beeinflusst haben. Entsprechend erstaunt es nicht, dass das fachdidaktische Wissen zu den anderen beiden Kategorien in geringerer Qualität zu beobachten war. Dieser Befund ist mit den von Zierer et al. (2015) berichteten Ergebnissen deckungsgleich und entspricht Erkenntnissen von John (2006). Es scheint, dass auch in der vorliegenden Studie der vorgegebene Planungsraster die angehenden Lehrpersonen bei ihrer Planungstätigkeit geleitet und eingeschränkt hat und dies zu entsprechend weniger Planungsüberlegungen geführt hat.

Ergebnisse aus der anschliessend erfolgten Untersuchung der Planungsunterlagen der Interventions- und Kontrollgruppe haben gezeigt, dass das fachdidaktische Wissen der Kontrollgruppe zur Planung eines Experimentalunterrichts im Sinne der Erkenntnisgewinnung im Vergleich zur Interventionsgruppe signifikant besser war, wobei der Unterschied eine schwache bis mittlere Effektstärke aufwies. Ein signifikanter Unterschied war zudem auch für die Dimension der «Analyse» zu verzeichnen. Diese Befunde könnten darauf hinweisen, dass die Planungsüberlegungen der Kontrollgruppe, die sich nur auf den Planungsraster gestützt hat, stärker durch das Instrument beeinflusst worden sind. Dass die Kategorie «Schülervorstellungen» ebenfalls signifikant in besserer Qualität erfasst werden konnte, ist auf den ersten Blick jedoch erstaunlich. Vor dem Hintergrund, dass die Interventionsgruppe zusätzlich mit den Kernaspekten gearbeitet und diese als Grundstruktur für die Gesprächsführung eingesetzt hat (vgl. Kap. 1.3), stellt sich die Frage, ob nicht ein *Cognitive Overload* die Qualität fachdidaktischen Wissens in den Unterrichtsplanungen, -materialien und videografierten Planungsgesprächen im Vorfeld beeinflusst hat (cf. Sweller, van Merriënboer & Paas, 2019). Die grundlegende Annahme der *Cognitive Load Theory* ist, dass der Wissenserwerb auf der Integration neuer Informationen aus dem Arbeitsgedächtnis in vorhandene Schemata des Langzeitgedächtnisses beruht und die Informationsaufnahme durch die beschränkte Kapazität des Arbeitsgedächtnisses unnötig belastet werden kann (Wilhelm & Kunina-Habenicht, 2015). Dabei kann die kognitive Belastung beim Lernen und Problemlösen durch die Komplexität des Inhaltes und das Vorwissen der Person (*intrinsic cognitive load*), durch das dargebotene Lernmaterial (*extraneous cognitive load*) oder aber durch die

Ressourcen im Arbeitsgedächtnis (*germane cognitive load*) hervorgerufen werden (Sweller et al., 2019). Entsprechend können Instruktionsprozesse, die auf verschiedenen interaktiven Elementen beruhen, für den Aufbau von Expertise bei Novizen eher kontraproduktiv sein (Sweller et al., 2019), was die Unterschiede zwischen der Interventions- und Kontrollgruppe hier erklären würde. Eine weitere Erklärung könnte allerdings auch in der Bildung der beiden Untersuchungsgruppen zu finden sein, die aufgrund des quasi-experimentellen Designs nicht vollständig randomisiert zusammengesetzt worden sind. Allerdings haben die Ergebnisse der ersten Tests (FDW t1, t2) und des Fachwissens (t1) keine signifikanten Unterschiede zwischen diesen beiden Gruppen ergeben.

Die hier diskutierten Befunde zu Vorkommen und Qualität experimentellen Handelns in Unterrichtsplanungen, -materialien und videografierten Planungsgesprächen verdeutlichen, dass mit der Analyse von schriftlichen Unterrichtsplanungen Aussagen über die Qualität von Unterrichtskonzeptionen als Antizipation einer Handlungssituation möglich sind (vgl. Davies & Rogers, 2000; König et al., 2015; Seel, 2011; Tillema, 2009). Sie sollen im nächstfolgenden Kapitel (Kap. II, 4.2) in Bezug zu den Ergebnissen der Professionswissenstests gesetzt, besprochen und auf der Grundlage von bereits bekannten Befunden eingeordnet werden.

4.2 Befunde zu möglichen Zusammenhängen zwischen der Planungsqualität und den Ergebnissen der Professionswissenstests (Forschungsanliegen 2b)

Befunde zum zweiten Forschungsanliegen 2b zeigen, dass das getestete Professionswissen und das in den Planungsunterlagen erfasste fachdidaktische Wissen zusammenzuhängen scheinen. Diejenigen Studierenden, die bessere Ergebnisse im fachdidaktischen Professionswissenstest vorweisen konnten, haben auch einen qualitativ besseren Unterricht zum Experimentieren im Sinne der Erkenntnisgewinnung gestaltet. Entsprechende Zusammenhänge zeigen sich auch zwischen den Testresultaten und den beiden Planungsdimensionen «Analyse» und «Konstruktion». Diese Ergebnisse sind absolut kongruent mit den Erkenntnissen aus der aktuellen Studie von Stender et al. (2017) und vergleichbar mit Befunden aus anderen Studien (Lange et al., 2012; Olszewski, 2010; Park et al., 2011; Smit, Rietz et al., 2017). Die Korrelationen mit Effektstärken im mittleren bis hohen Bereich stützen darüberhinaus die Validität des für die Analyse der Planungsmaterialien verwendeten Ratingmanuals (Kap. I, 4.1).

Die ermittelten Korrelationen zwischen dem getesteten Fachwissen und dem in den Planungsunterlagen erfassten fachdidaktischen Wissen sind allerdings eher erstaunlich, da mit

dem verwendeten Fachwissenstest das Wissen zur «Visuellen Wahrnehmung» geprüft worden ist und nicht das Fachwissen zum Experimentieren im Sinne der Erkenntnisgewinnung. Es wäre durchaus denkbar, dass ein grundsätzlich besseres Fachwissen zum biologischen Fachinhalt, der als fachlicher Aufhänger für die Planung gewählt worden ist, die Qualität der Planungen beeinflussen kann. Gemäss dem *Expertise Reversal Effect*, einem Instruktioneffekt der Cognitive Load Theory, verschwindet die Limitierung des Arbeitsgedächtnisses vor dem Hintergrund, dass bereits gelernte Informationen aus dem Langzeitgedächtnis zur Verfügung stehen (Kalyuga, Rikers & Paas, 2012). Entsprechend könnte hier argumentiert werden, dass eine gute fachwissenschaftliche Grundlage Raum für die Auseinandersetzung mit den fachdidaktischen Fragen schafft (cf. Baumert & Kunter, 2011). Diese Argumentation wäre allerdings nur dann plausibel, wenn auch das Fachwissen zum Experimentieren im Sinne der Erkenntnisgewinnung korrelative Zusammenhänge mit dem in den Planungen erfassten fachdidaktischen Wissen zeigen würde. Da hierzu Testwerte fehlen und keine Aussagen gemacht werden können, aber auch keine signifikanten Zusammenhänge zwischen dem getesteten fachdidaktischen Wissen und dem Fachwissen ermittelt werden konnten, was für zwei nicht kongruente Testkonstrukte sprechen könnte, kann dieses Ergebnis nur mit grosser Vorsicht entsprechend interpretiert werden.

Rückblickend lässt sich festhalten, dass die kombinierte Erfassung des Professionswissens mit Hilfe des im Projekt entwickelten Tests und dem Ratingmanual zur Analyse der erstellten Unterrichtsplanungen, einen breiteren Einblick in die Planungskompetenzen der Studierenden gewährt hat. Mit diesem Setting konnte der hierzu vielfach geäusserten Kritik entgegengewirkt werden, dass Wissenstests keine umfassenden Kompetenzmessungen darstellen (Oser, Heinzer & Salzmann, 2010, Hasse, Joachim, Bögeholz & Hammann, 2014) und es daher einer Betrachtung der Handlungsebene bedarf, um das getestete Wissen einzuordnen (Vogelsang & Reinhold, 2013). Die hier diskutierten Ergebnisse basieren allerdings auf einer Untersuchung, die zur Hauptsache an einer Stichprobe von 70 Studierenden durchgeführt worden ist. Vor diesem Hintergrund müssen deshalb auch die Aussagekraft und Limitationen dieser Teilstudie II im nächsten Kapitel erläutert werden.

4.3 Grenzen und Aussagekraft der Teilstudie II

Obgleich aufgrund der umfassenderen Kompetenzmessung in Teilstudie II Aussagen zur Planungskompetenz angehender Lehrpersonen hinsichtlich der Gestaltung eines Unterrichts zum experimentellen Handeln möglich sind, haben sich Limitationen in Bezug auf gewisse Erhebungsinstrumente, aber auch hinsichtlich des Designs gezeigt, die hier noch erläutert werden sollen.

Mit den im Projekt KUBeX entwickelten und eingesetzten Tests wurden das fachdidaktische Wissen zur Planung von Unterricht zum experimentellen Handeln, die Beliefs und das Fachwissen der Studierenden zu verschiedenen Messzeitpunkten erhoben. Wie sich herausgestellt hat, war nur der Fachwissenstest zum experimentellen Handeln nicht reliabel. Direkte Zusammenhänge zwischen der Planungsqualität und dem themenspezifischen Fachwissen zum experimentellen Handeln konnten daher nicht ermittelt werden (cf. Stender, 2014; Stender et al., 2015, 2017). Für eine bessere Einordnung der signifikanten Korrelation zwischen dem erhobenen Fachwissen zur «Visuellen Wahrnehmung» und der Qualität der Planungen wäre dies hilfreich gewesen. Es muss allerdings hier ergänzend angemerkt werden, dass die berechneten Korrelationen keine Aussagen zu Ursachen oder Wirkungen erlauben und es daher hierzu noch weitere Auswertungsverfahren bräuchte. Im Projekt KUBeX wurde zudem aus testökonomischen Gründen kein Test zur Motivation eingesetzt. Einflüsse der Motivation auf die Planungsqualität, die Stender nachweisen konnte (Stender, 2014; Stender et al., 2015, 2017), hätten noch weitere Aufschlüsse zu den erfassten Ergebnissen liefern können.

Mit dem im Projekt KUBeX entwickelten Ratingmanual konnten alle Hinweise zum experimentellen Handeln in den Planungsmaterialien valide erfasst werden. Aussagen über die Qualität der Unterrichtskonzeptionen waren mit diesem Verfahren prinzipiell möglich. Allerdings muss hier eingewendet werden, dass der Planungsprozess vielschichtig ist und von den Akteuren meist nicht vollständig verbalisiert wird (Seel, 2011). Entsprechend gewährt die Analyse von Unterrichtsplanungen mit Hilfe von Ratingmanuals nur einen limitierten Einblick in das fachdidaktische Wissen von (angehenden) Lehrpersonen, insbesondere auch, weil es nie möglich ist, alle Wissensfacetten vollständig zu erfassen (Park & Suh, 2015). Zu berücksichtigen ist auch, dass es sich hier um eine reine Momentaufnahme handelt und eine Vergleichsmöglichkeit mit einer Baseline fehlt. Der Schluss, dass die angehenden Lehrpersonen der untersuchten Stichprobe Mühe bekundet haben, einen Unterricht zum experimentellen Handeln zu planen, sollte daher vor diesem Hintergrund entsprechend eingeordnet werden. Mit Hilfe zusätzlicher reflexiver Verfahren wie Interviews oder *stimulated recalls* hätten die Erkenntnisse aus den Ratings noch breiter abgestützt werden können. Auch stellen solche Reflexionsanlässe eine wichtige Voraussetzung dar für die iterative Weiterentwicklung von Unterrichtsplanungen und für die Festigung der fachdidaktischen Handlungspläne (Stender, 2014, S. 39; Stender et al., 2017). Die Vorgaben in der Vignette haben vermutlich zusätzlich noch dazu beigetragen, dass Hinweise zum Lehrplan wie auch zu Individualisierungsmaßnahmen in den Planungsunterlagen der Studierenden gefehlt haben. Entsprechende Anpassungen wären hier vorzunehmen. Ausserdem sei abschliessend noch darauf hingewiesen, dass die Validierung des Ratingmanuals und

die anschliessend erfolgte qualitative Auswertung der Planungsmaterialien im Sinne einer Pilotierung auf der gleichen Datenbasis erfolgt sind. Eine erneute und im Sinne von Messick (1995, Kap. I, 3.2.5) noch breiter abgestützte Validierung des Ratingmanuals an einer anderen, wenn möglich grösseren Stichprobe, müsste in einem nächsten Schritt daher zwingend vorgenommen werden.

Auch wenn die Teilstudie II erste Einblicke in das Planungshandeln der Studierenden ermöglicht hat, so bleibt die Frage nach wie vor offen, ob die Qualität der erstellten Planungen auch ein guter Prädiktor für das konkrete Unterrichtshandeln, die Unterrichtsqualität und damit auch die Lernergebnisse der Schülerinnen und Schüler darstellt, wie dies in anderen Studien bereits nachgewiesen werden konnte (Baumert et al., 2010; Hill et al., 2005; Olszewski, 2010). Vor dem Hintergrund dieser noch offenen Frage sollen im letzten Kapitel (Kap. II, 5) die beiden Teilstudien zusammengeführt werden, um dann einen Ausblick im Sinne möglicher Implikationen für weiterführende Forschungsprojekte sowie für die Lehrpersonenbildung und den Professionalisierungsprozess von angehenden Lehrpersonen zur Planung von Unterricht zum experimentellen Handeln zu geben.

5 Synopse beider Teilstudien und Ausblick

Die Handlungskompetenz von Lehrpersonen wird als zentrale Handlungsressource oder *knowledge base of teaching* angesehen (Harms & Riese, 2018; Verloop et al., 2001). Wissen und Können von Lehrpersonen – neben den Einstellungen, Überzeugungen, motivationalen Orientierungen und selbstregulativen Fähigkeiten – stellen im Sinne des Expertenparadigmas die zentralen Komponenten ihrer Professionalität dar (Harms & Riese, 2018; Jüttner & Neuhaus, 2013). Sie gelten als wichtige Voraussetzungen für erfolgreichen Unterricht (Alonzo & Kim, 2016; Baumert & Kunter, 2006, 2011a; Besser & Krauss, 2009; Kirschner et al., 2016; Künsting et al., 2009; Pauli & Reusser, 2009; Wagner, 2014; Olga Zlatkin-Troitschanskaia et al., 2009b) und wesentlichen Einflussfaktor für die Kompetenzentwicklung von Schülerinnen und Schülern (u.a. Kunter et al., 2013; Lipowsky, 2009). Dabei stehen vor allem kognitive und affektive Dispositionen von Lehrpersonen als Voraussetzung für kompetentes Handeln im Unterricht im Fokus (Baumert & Kunter, 2011b; Ohle et al., 2011; Riese, 2009). Diese aufzubauen ist Aufgabe der Lehrpersonenbildung.

Mit den neuen Bildungsstandards (EDK, 2011; KMK, 2005) sind (angehende) Lehrpersonen verpflichtet, Kompetenzen im Bereich des Experimentierens im Sinne der Erkenntnisgewinnung zu fördern. Bisherige Studien zeigen, dass es Lehrpersonen dafür u.a. an fachlichem Wissen und Verständnis zum Konzept *Scientific Inquiry*, aber auch an fachdidaktischem Wissen mangelt (e.g. Gyllenpalm & Wickman, 2011a). Bisher sind Untersuchungen zur professionellen Handlungskompetenz hauptsächlich auf der Basis von Unterrichtsanalysen mit Hilfe von *Video Surveys* (z.B. Börlin, 2012; Seidel et al., 2006), aber auch Verfahren wie *video-stimulated recalls* und videografierten Unterrichtsbesprechungen (z.B. Pauli & Reusser, 2006; Riegel, 2013) sowie Tests erfolgt (e.g. Tepner et al., 2012; Bölsterli et al. 2010). Kompetenzmessungen und -modellierungen mit Hilfe von Unterrichtsplanungen haben im Gegensatz dazu noch wenig Tradition (von Aufschnaiter & Blömeke, 2010). Die Befundlage zu den bisher durchgeführten standardisierten Analysen von Unterrichtsplanungen zeigt deshalb auch, dass keine konsistente Beschreibung des Planungshandelns und damit der Planungskompetenz von Lehrpersonen vorliegt (Vogelsang und Riese, 2017; Weingarten und van Ackeren, 2017). Auch bleibt nach wie vor unklar, wie sich die Planungskompetenz tatsächlich operationalisieren lässt (Zierer et al., 2015b). Beschreibungen der entsprechenden Expertise zum Planen von Unterricht zum experimentellen Handeln als Fachinhalt wie auch Erhebungsinstrumente, mit deren Hilfe eine standardisierte Erfassung der Qualität der Planungsperformanz in Bezug auf das beobachtbare

fachdidaktische Wissen erfolgt, sind daher selten und inkonsistent (Hasse et al., 2014, Voggelang und Riese, 2017; Weingarten und van Ackeren, 2017). Da bisher wenig bekannt ist, über welches fachdidaktische Wissen angehende Lehrpersonen zum Experimentieren im Sinne der Erkenntnisgewinnung tatsächlich verfügen und in welcher Form sich dieses, in Unterrichtsplanungen angewendete Wissen, tatsächlich erfassen lässt, hat sich die vorliegende Studie diesem Anliegen gewidmet. Im folgenden Kapitel 5.1 werden die Ergebnisse aus beiden Teilstudien zusammengefasst und eingeordnet.

5.1 Synopse der beiden Teilstudien I & II

Die vorliegende Studie hat zwei Forschungsanliegen verfolgt. Im Rahmen des ersten Forschungsanliegens wurde das im Projekt KUBeX entwickelte Ratinginstrument zur Beurteilung von Unterrichtsplanungen, die das Ziel haben, Schülerinnen und Schüler Methoden der Erkenntnisgewinnung erwerben zu lassen, empirisch geprüft und soweit möglich validiert. Auch wurde das Ratinginstrument weiter in Bezug auf die beiden zentralen Dimensionen des Modells der «Didaktischen Rekonstruktion» («Analyse», «Konstruktion») analysiert und empirisch geprüft. Auf der Grundlage der in diesem Zusammenhang erfolgten Analysen, wurden im Rahmen des zweiten Forschungsanliegens erste Antworten auf die Frage, ob angehende Lehrpersonen Unterricht zum Experimentieren im Sinne der Erkenntnisgewinnung planen und wenn ja, welche Qualität fachdidaktischen Wissens sich in den Planungen zeigt, ermittelt. Diese Ergebnisse wurden zudem in Bezug zu weiteren, im Projekt KUBeX erfassten Faktoren, wie z.B. dem getesteten fachdidaktischen Wissen, untersucht und werden im folgenden Kapitel zusammengefasst.

5.1.1 Zentrale Erkenntnisse der vorliegenden Studie (Teilstudie I & II)

Zentrale Erkenntnisse der hier mittels ESEM im Programm MPlus Version 8.0 (Muthén & Muthén, 1998-2017) erfolgten explorativen Untersuchungen zum ersten Forschungsanliegen haben ergeben, dass das Ratingmanual auf der Basis der untersuchten Stichprobe faktoriell valide ist. Es hat sich gezeigt, dass das 2-Faktorenmodell signifikant besser auf die Daten passt als das 1-Faktormodell. Zudem erwies sich die im 2-Faktorenmodell auf der Basis der Faktorladungen erfolgte Zuteilung der Indikatoren zu den Faktoren inhaltlich als vereinbar mit den Wissensdimensionen «Analyse» und «Konstruktion» aus dem Modell der «Didaktischen Rekonstruktion». Daraus kann die Hypothese abgeleitet werden: Wenn Unterricht im Sinne der Tiefenstruktur auf Lernprozessebene geplant wird, ist der Planungsprozess durch zwei Phasen gekennzeichnet, deren Tätigkeiten jeweils eng in Zusammenhang zueinander stehen. Die Ergebnisse sprechen dafür, dass in der Phase der Analyse, bei der es um die Klärung der Kompetenzen der Lernenden, die Klärung der Kompetenz-

ansprüche im Fach und die Vermittlungsabsichten geht, wichtige Grundlagen geschaffen werden für die Konstruktion qualitätsvoller Lerngelegenheiten. Die Konstruktion von Unterricht kann somit nur unter Rückgriff auf die Analyse rekursiv erfolgen. Entsprechend kommt der Analysephase eine wichtige Bedeutung zu. In gewisser Weise sind hier auch Parallelen zum 2-dimensionalen Konstrukt von Gess-Newsome et al. (2017) erkennbar. Allerdings werden im 2-Faktorenmodell der vorliegenden Studie die Kategorien «Schülervorstellungen» und «Reflexivität» eher mit dem Fachwissen und weniger mit dem pädagogischen Wissen in Verbindung gebracht. Das lässt sich damit begründen, dass die Kenntnisse von und der Umgang mit Schülervorstellungen, aber auch die Reflexion zu den Ergebnissen und den Fachbegriffen enger mit dem Fachwissen verknüpft sind.

Befunde zum zweiten Forschungsanliegen, die mit Hilfe des vorangehend validierten Instruments in SPSS (IBM Corp., 2016) ermittelt worden sind, zeigen, dass nur bei 70 aller 119 Planungsmaterialien Hinweise zum Experimentieren im Sinne der Erkenntnisgewinnung zumindest ansatzweise festgestellt werden konnten. Diese Teilstichprobe ($n = 70$) unterscheidet sich ausserdem signifikant von der Reststichprobe ($n = 49$) und weist unterschiedliche Kompetenzausprägungen der themenspezifischen Planungskompetenz auf, wobei tiefere Kompetenzausprägungen vergleichsweise häufiger sind. Insgesamt deutet dies darauf hin, dass die Studierenden der untersuchten Stichprobe Schwierigkeiten bekundet haben, einen entsprechenden Unterricht zu gestalten und ihnen trotz der Impulsveranstaltung I möglicherweise die nötigen Fähigkeiten und Fertigkeiten dazu gefehlt haben (cf. Fazio et al., 2010). Die ermittelten Korrelationen zwischen dem getesteten fachdidaktischen Wissen und dem in den Planungsmaterialien erfassten fachdidaktischen Wissen stützen diese Interpretation. Allerdings könnten auch die in Kapitel I, 5.3 bzw. Kapitel II, 4.2 angesprochene Konfundierung mit anderen Planungsaufträgen oder aber auch die Motivation die Planungsqualität beeinflusst haben. Es hat sich weiter herausgestellt, dass in den Planungsunterlagen das fachdidaktische Wissen zu den Kategorien «Fachliche Klärung», «Unterrichtsziele» und «Didaktische Strukturierung» im Vergleich zu den Kategorien «Schülervorstellungen» und «Reflexivität» in besserer Qualität erfasst worden ist. Ein Vergleich mit den Rubriken im verwendeten Planungsraster zeigt, dass die drei Kategorien, die in den Unterrichtsplanungen qualitativ besser umgesetzt worden sind, im Planungsraster explizit aufgeführt waren (John, 2006; Zierer, Werner et al., 2015). Entsprechende Auswirkungen haben sich auch bei den beiden Dimensionen des Modells der «Didaktischen Rekonstruktion», «Analyse» und «Konstruktion», ergeben. Die Qualität des beobachteten fachdidaktischen Wissens der Dimension «Konstruktion» ist im Vergleich zur Dimension «Analyse» in den Planungsmaterialien in besserer Qualität umgesetzt worden. Erstaunlicherweise konnte in den Planungsmaterialien der Kontrollgruppe im Vergleich zur Inter-

ventionsgruppe eine signifikant bessere Qualität fachdidaktischen Wissens zur Planung von Unterricht zum experimentellen Handeln festgestellt werden. Die Effektstärken sind allerdings nur schwach bis mittel. Die signifikanten Unterschiede können vorwiegend auf Indikatoren der Dimension «Analyse» zurückgeführt werden. Dabei weisen Indikatoren der Kategorien «Unterrichtsziele», «Fachliche Klärung» und «Schülervorstellungen» in den Planungsmaterialien der Kontrollgruppe eine signifikant bessere Qualität auf. Auch hier könnten die Planungsüberlegungen der Kontrollgruppe, die sich nur auf den Planungsraster gestützt hat, stärker durch das Instrument beeinflusst worden sein. Der Befund, dass die Kategorie «Schülervorstellungen» bei der Kontrollgruppe ebenfalls signifikant in besserer Qualität erfasst werden konnte, ist eher erstaunlich. Er könnte aber darauf zurückzuführen sein, dass die Interventionsgruppe zusätzlich mit den Kernaspekten gearbeitet und diese als Grundstruktur für die Gesprächsführung eingesetzt hat. Der Einsatz zweier Instrumente (Planungsraster, Kernaspekte) hat möglicherweise zu einer kognitiven Belastung bei den Studierenden geführt (Sweller et al., 2019), was die Unterschiede zwischen der Interventions- und Kontrollgruppe hier erklären könnte.

Die in der Folge ermittelten Zusammenhänge zwischen dem getesteten deklarativen fachdidaktischen Wissen zur Planung von Unterricht zum experimentellen Handeln und der Qualität der gesamten Planungsunterlagen sind signifikant. Daraus kann abgeleitet werden, dass diejenigen Studierenden, die im fachdidaktischen Professionswissenstest besser abgeschnitten haben, auch einen qualitativ besseren Unterricht zum Experimentieren im Sinne der Erkenntnisgewinnung gestaltet haben. Entsprechende Korrelationen konnten auch zwischen den Testresultaten und den beiden Planungsdimensionen «Analyse» und «Konstruktion» nachgewiesen werden. Diese Ergebnisse sind absolut kongruent mit den Erkenntnissen aus der aktuellen Studie von Stender et al. (2017) und vergleichbar mit Befunden aus anderen Studien (Lange et al., 2012; Olszewski, 2010; Park et al., 2011; Smit, Rietz et al., 2017). Die Korrelationen mit Effektstärken im mittleren bis hohen Bereich stützen darüber hinaus die Validität des für die Analyse der Planungsmaterialien verwendeten Ratingmanuals (Kap. 4.1. der Teilstudie I). Ein signifikanter Zusammenhang zwischen dem getesteten deklarativen Fachwissen und der latenten Variablen «Experimentelles Handeln» bzw. den beiden Dimensionen «Analyse» und «Konstruktion» konnte ebenfalls ermittelt werden. Dies ist allerdings erstaunlich, da mit dem verwendeten Fachwissenstest das Wissen zur «Visuellen Wahrnehmung» geprüft worden ist und nicht das Fachwissen zum Experimentieren im Sinne der Erkenntnisgewinnung. Grundsätzlich ist denkbar, dass ein besseres Fachwissen zum biologischen Fachinhalt, der als fachlicher Aufhänger für die Planung gewählt worden ist, die Qualität der Planungen beeinflussen kann. Gemäss dem *Expertise Reversal Effect*, einem Instruktionseffekt der *Cognitive Load Theory*, verschwindet die

Limitierung des Arbeitsgedächtnisses vor dem Hintergrund, dass bereits gelernte Informationen aus dem Langzeitgedächtnis zur Verfügung stehen (Kalyuga et al., 2012). Entsprechend könnte eine gute fachwissenschaftliche Grundlage Raum für die Auseinandersetzung mit den fachdidaktischen Fragen schaffen (cf. auch Baumert & Kunter (2011). Diese Argumentation wäre allerdings nur dann plausibel, wenn auch das Fachwissen zum Experimentieren im Sinne der Erkenntnisgewinnung korrelative Zusammenhänge mit dem in den Planungen erfassten fachdidaktischen Wissen zeigen würde. Da hierzu Testwerte fehlen und damit keine Aussagen gemacht werden können, sind diese Ergebnisse mit Vorsicht zu interpretieren.

Rückblickend lässt sich festhalten, dass in der vorliegenden Studie in Bezug auf (a) die Erfassung des fachdidaktischen Wissens zur Planung von Unterricht zum experimentellen Handeln, (b) die Gestaltung eines entsprechenden Unterrichts und (c) hinsichtlich der für die Planung eingesetzten Instrumente aufschlussreiche Erkenntnisse gewonnen werden konnten. Diese basieren allerdings mehrheitlich auf einer Stichprobe von 70 Studierenden der Sekundarstufe I. Die Validierung des Ratingmanuals und die Qualitätsanalyse der Planungsmaterialien und der videografierten Planungsgespräche sind zudem auch an der gleichen Stichprobe erfolgt. Entsprechend sind Limitierungen zu verzeichnen, die die Aussagekraft beeinflussen. Diese werden im nächsten Kapitel (Kap. II, 5.1.2) kurz zusammengefasst.

5.1.2 Grenzen und Aussagekraft der vorliegenden Studie

Die Kompetenzerfassung in der vorliegenden Studie ist durch ein vergleichsweise aufwändiges Beobachtungsverfahren erfolgt. Die Stichprobengrösse ist daher, aber auch bedingt durch die jeweiligen Ausbildungssituationen an den vier Pädagogischen Hochschulen mit $N = 119$ vergleichsweise klein. Die Bildung der Interventions- und Kontrollgruppe erfolgte deshalb nur mehrheitlich randomisiert. Obwohl die Studie direkt in eine Lehrveranstaltung mit Praxisbezug eingebettet war, musste der Planungsauftrag zur Gewährleistung der Vergleichbarkeit ausserdem auf der Basis einer Vignette erfolgen. Damit ging die Authentizität der Planungssituation verloren und eine mögliche Konfundierung mit weiteren Planungsaufträgen aus anderen Modulen konnte nicht vermieden werden.

Für die Analyse der Unterrichtsplanungen, -materialien und videografierten Planungsgespräche erwies sich das Ratingmanual als faktoriell valide. Allerdings unterschied das Instrument zu wenig eindeutig zwischen dem Fehlen der Planungskompetenz zum experimentellen Handeln und dem Vorhandensein in verschiedenen Qualitätsstufen, weshalb es hier weiterer Codes bedürfte. Entsprechend zeigte sich dies auch in einem erheblichen Bodeneffekt. Die meisten Analysen mussten deshalb auf der Basis der Teilstichprobe

$n = 70$ durchgeführt werden, die das experimentelle Handeln zumindest ansatzweise in den Planungsmaterialien umgesetzt haben. Auch die faktorielle Validität wurde nur an dieser Teilstichprobe ($n = 70$) geprüft. Dabei hat sich gezeigt, dass mit 14 der 17 Indikatoren, das Strukturmodell zur Erfassung des fachdidaktischen Wissens zur Planung von Unterricht faktoriell validiert werden konnte. Die aus den Befunden abgeleitete Hypothese, dass der Planungsprozess von Unterricht im Sinne der Tiefenstruktur auf Lernprozessebene durch zwei Phasen gekennzeichnet ist, deren Tätigkeiten jeweils eng in Zusammenhang zueinander stehen, müsste in einem nächsten Schritt an einer anderen Stichprobe tatsächlich geprüft werden. Ebenso müssten alle Indikatoren erneut auf deren Tauglichkeit für die Validierung überprüft werden. Auch würde es sich lohnen, die Messgenauigkeit des Ratingmanuals mit einer Generalisierbarkeitsstudie zu ermitteln (Cronbach et al., 1963), da mit diesem Verfahren mehrere Fehlerquellen gleichzeitig untersucht werden könnten (cf. Park & Suh, 2015).

Die Qualitätsanalyse der Planungsmaterialien ist ebenfalls an der Teilstichprobe ($n = 70$) im Sinne einer explorativen Herangehensweise erfolgt und sollte vor dem Hintergrund, dass eine Baseline fehlt, im Sinne einer Momentaufnahme eingeordnet werden. Die so ermittelten Befunde zum fachdidaktischen Wissen zur Gestaltung von Unterricht zum experimentellen Handeln müssen ausserdem aufgrund der Tatsache, dass Planungsprozesse vielschichtig sind und von den Akteuren auch nicht immer vollständig verbalisiert oder beschrieben werden, in diesem Sinn eingestuft werden. Auch hat sich herausgestellt, dass die Vorgaben in der Vignette vermutlich dazu geführt haben, dass Hinweise zum Lehrplan wie auch zu Individualisierungsmassnahmen in den Planungsunterlagen der Studierenden gefehlt haben. Entsprechende Anpassungen wären hier nötig.

Die im Projekt KUBeX weiterführenden Tests haben die Überprüfung möglicher Zusammenhänge zwischen der Qualität des fachdidaktischen Wissens in den Planungsmaterialien und den Tests ermöglicht. Allerdings konnten keine Bezüge zum Fachwissenstest zum experimentellen Handeln aufgrund fehlender Reliabilität dieser Tests gemacht werden. Zudem konnte das Planungsergebnis nicht in Bezug auf die Motivation geprüft werden, da diese affektive Disposition im Projekt KUBeX nicht erfasst worden ist.

Für weiterführende Untersuchungen müssten alle hier beschriebenen Aspekte mitberücksichtigt, verändert und weiterentwickelt werden. Vor diesem Hintergrund soll das nächste Kapitel (Kap. II, 5.2) einen Ausblick zu weiterführenden Forschungsvorhaben bieten und mögliche Implikationen für die Professionalisierung der Lehrpersonenbildung aufzeigen.

5.2 Ausblick und mögliche Implikationen für die Lehrpersonenbildung

Die in der vorliegenden Studie gewonnenen empirischen Befunde haben einen ersten Einblick in die Kompetenzausprägungen der Studierenden der vier Pädagogischen Hochschulen zum fachdidaktischen Wissen zur Planung von Unterricht zum experimentellen Handeln gewährt. Die Erfassung des Professionswissens mit Hilfe der im Projekt entwickelten Tests (Disposition) in Kombination mit dem Ratingmanual zur Analyse der erstellten Planungsmaterialien hinsichtlich des handlungsnäheren fachdidaktischen Wissens (Performanz), hat dabei einen breiteren Einblick in die Kompetenzausprägungen der Studierenden ermöglicht. Allerdings können aus diesen Befunden keinerlei Voraussagen dazu gemacht werden, ob und wie sich die erfasste Planungsqualität auch auf den Unterricht auswirkt und ob diese die Schülerleistungen beeinflusst. Ausgehend von dieser Frage befassen sich die nächsten beiden Abschnitte (I.) mit einem Ausblick auf künftige, weiterführende Forschungsvorhaben und (II.) mit Implikationen für die Professionalisierung angehender Lehrpersonen und die Lehrpersonenausbildung in Bezug auf das experimentelle Handeln im Sinne der Erkenntnisgewinnung.

I. Ausblick auf weiterführende Forschungsvorhaben

Im Kontext der Lehrpersonenbildung und der Professionalisierungsforschung stellt das Planungshandeln eine wichtige Schlüsselstelle des Unterrichtens dar (Zierer et al., 2015a). Das Erstellen von Unterrichtsplanungen spielt für die Entwicklung kompetenten Handelns und in diesem Zusammenhang auch von PCK eine wesentliche Rolle (Ball & Bass, 2000; Baylor, 2002). Die Forschung hinsichtlich der professionellen Handlungskompetenz von (angehenden) Lehrpersonen wurde in den vergangenen Jahren intensiv vorangetrieben, wobei sich gerade in den naturwissenschaftlichen Fachdidaktiken Biologie, Chemie und Physik vergleichsweise wenige empirische Studien mit diesem Thema befassen. Entsprechend fehlen hierzu nach wie vor konsistente Beschreibungen, ob und wie sich das Professionswissen der Lehrperson ganz grundsätzlich über ihr konkretes Unterrichtshandeln auf die Unterrichtsqualität und die Lernergebnisse der Schülerinnen und Schüler auswirkt (Borowski et al., 2010; Kirschner et al., 2017, Kap.II, 1.1.2). Insbesondere Fragestellungen, die den Zusammenhang zwischen Unterrichtsplanung und konkreter Unterrichtsdurchführung berücksichtigen, sind vor dem Hintergrund der Lehrpersonenprofessionalisierung bisher eher vernachlässigt worden (Seel, 2011, S. 32). Hinweise dazu sind in einigen wenigen Studien mit allgemeindidaktischem (König et al., 2015; Seel, 1997), aber auch mit fachdidaktischem Fokus zu finden (Anthofer, 2017; Gess-Newsome et al., 2017), wobei Studien

zum experimentellen Handeln im Sinne der Erkenntnisgewinnung fehlen. In der allgemein-didaktisch ausgerichteten Studie von Seel, 1997 (S. 264) konnte zwischen den von Studierenden im Planungsgespräch geäußerten Intentionen und dem realisierten Unterricht kein Zusammenhang festgestellt werden. Das Augenmerk dieser Studie lag allerdings hauptsächlich auf der Organisation und Strukturierung von Unterricht. König, Buchholtz und Dohmen (2015) fanden demgegenüber erste Hinweise, wonach Schülerinnen und Schüler in der Unterrichtsstunde einen stärker binnendifferenzierenden Unterricht wahrnehmen, wenn bereits in der Unterrichtsplanung die didaktische Adaptivität höher ausgefallen ist. In der naturwissenschaftsdidaktisch ausgerichteten Studie von Gess-Newsome et al. (2017) konnte der erhoffte Zusammenhang zwischen dem in Planungsinterviews erfassten PCK, dem Unterricht und Schülerleistungen empirisch nicht nachgewiesen werden. Die Frage adäquater Messinstrumente wurde hier u.a. diskutiert. Eine aktuelle Studie von Anthofer (2017) zum Planen und Unterrichten des Themas «Experimentieren bei Stofftrennungen» konnte hingegen über die Analyse unterrichtlichen Handelns und nachgeschalteter Interviews Zusammenhänge zwischen der Qualität der Unterrichtsvorbereitungen und dem unterrichtlichen Handeln von Studierenden feststellen. Die Qualität der Unterrichtsplanungen wurde hier allerdings nur indirekt erfasst. Die Frage, ob die tatsächlich erfassbare Qualität von Unterrichtsplanungen auch ein Prädiktor für einen qualitätsvollen Unterricht, d.h. guten und effektiven Unterricht (Berliner, 2005, S. 207, Kap. I, 2.6), darstellt, bleibt vor dem Hintergrund der hier geschilderten Befunde nach wie vor ungeklärt, insbesondere auch in Bezug auf die Gestaltung von Unterricht zum experimentellen Handeln im Sinne der Erkenntnisgewinnung.

Für die Ausbildung und Professionalisierung angehender Naturwissenschaftslehrpersonen haben sich Transformationsmodelle, wie das Planungsmodell von Stender (2014, 2015, 2017), als nützlich erwiesen. Sie implizieren, dass die Entwicklung des themenspezifischen Professionswissens einem Mechanismus unterliegt, der eine gute theoretische Grundlage für die Ausbildung darstellen könnte (cf. Kind, 2009). Das bereits 1987 entwickelte MPRA-Modell von Shulman (1987, Kap. I, 1.3) bildet die Entwicklung von PCK in Verbindung mit der unterrichtlichen Tätigkeit von Lehrpersonen in ähnlicher Weise ab und unterstreicht dabei die Relevanz der Reflexion in diesem Prozess (Fernandez, 2014, S. 81). Für den Aufbau der professionellen Handlungskompetenz und damit verbunden auch des themenspezifischen Professionswissens und Könnens angehender Lehrpersonen, sind demnach geeignete Reflexionsanlässe, die nachweislich zur Kompetenzentwicklung beitragen können, wichtig (e.g. Jacobs et al., 2008; van der Valk & Broekman, 1999). Die meisten Studien, die sich mit Unterrichtsreflexionen befassen, sind allerdings gekennzeichnet durch einen direkten Rückbezug auf einen zuvor gehaltenen und ebenfalls analysierten Unterricht

(Krammer et al., 2010; Mutton, Hagger & Burn, 2011c; Wyss, 2013). Eine der wenigen quantitativen Studien, die sich schwerpunktmässig mit Inhalten der vorbereitenden und nachbereitenden Unterrichtsreflexionen auseinandersetzt hat, stammt von Futter (2017). Die Autorin dieser Studie hat in einer Stichprobe mit 30 Unterrichts-Vorbesprechungen und 31 Unterrichts-Nachbesprechungen zwischen Praxislehrpersonen und den Studierenden der Sekundarstufe I abgrenzbare Sequenzen, sogenannte „potenzielle Lerngelegenheiten“, identifiziert. Dabei hat sie festgestellt, dass sich rund 69 Prozent der potenziellen Lerngelegenheiten mit der organisatorischen bzw. unterrichtsmethodischen Sichtstruktur des Unterrichts, 19 Prozent mit Fragen der Pädagogik (Schülerinnen und Schüler oder die Klasse als Ganzes) und nur 9 Prozent mit fachlichen oder fachdidaktischen Aspekten auseinandersetzen. Darüberhinaus wurden fachliche und fachdidaktische Aspekte – wenn überhaupt – signifikant häufiger in der Vorbereitung angesprochen als danach (Futter, 2017, S. 236). Demnach scheint dieser für die Gestaltung wirksamer Lerngelegenheiten wichtige Aspekt der professionellen Kompetenz von Lehrpersonen bei zukünftigen Lehrpersonen in Unterrichtsbesprechungen nicht im Vordergrund zu stehen. Vor dem Hintergrund, dass die Studierenden bereits zwei Drittel ihrer Ausbildung absolviert hatten, stuft die Autorin dies als nicht erwartungskonform ein (ebd. 2017, S. 237). Eine ähnliche Schlussfolgerung lässt sich auch aus der Arbeit von Wyss (2013) zu Unterrichtsreflexionen von Primarlehrpersonen ziehen. Ergebnisse aus den *Stimulated Recall*-Interviews dieser Studie ergaben, dass die Lehrpersonen in ihren Reflexionen vorrangig die «Qualität der Organisation», die «Motivierungsfähigkeit» und die «Klassenführung» fokussiert haben. Demnach scheinen Qualitätskriterien der Tiefenstruktur der Lernprozesse auch hier nicht thematisiert worden zu sein. Allerdings hat eine Analyse von Reflexionen zum Naturwissenschaftsunterricht auch gezeigt, wie herausfordernd es ist, mit Hilfe eines hoch inferneten Kodiermanuals fachdidaktische Tiefenstrukturmerkmale zu erfassen (Abels, 2011). Aktuell gehen deshalb mehrere Forschungsgruppen dieser zwischenzeitlich erkannten Herausforderung nach, ein valides Beurteilungsraster von Unterrichtsreflexionen zu fachdidaktischen Aspekten der Naturwissenschaften zu generieren (Kempin, Kulgemeyer, & Schecker, 2018; Vogelsang, Riese, Kulgemeyer & Borowski, 2017).

Ein Blick in die Literatur zeigt, dass es verschiedene Studien gibt zu Handlungskompetenzen in Bezug auf die Unterrichtsplanung (cf. Haas, 2005; König et al., 2015), die Reflexion (cf. Abels, 2011; Hatton & Smith, 1995) sowie das unterrichtliche Handeln (cf. Anthofer, 2017; Jatzwauk et al., 2008; Praetorius et al., 2014; Wadouh, Liu, Sandmann & Neuhaus, 2014; Wüsten, 2010), darunter auch Arbeiten mit naturwissenschaftlichem Hintergrund. Allerdings finden sich kaum Studien, die alle drei Handlungskompetenzen, d.h. das Planen von Unterricht, das unterrichtliche Handeln und die Reflexion über den Unterricht

gleichzeitig in Beziehung zueinander bringen. Zur Beantwortung der Frage, ob die tatsächlich erfassbare Qualität von Unterrichtsplanungen auch ein Prädiktor für einen qualitätsvollen Unterricht, d.h. guten und effektiven Unterricht (Berliner, 2005, S. 207, Kap. 2.6), darstellt, wäre eine solche Vorgehensweise notwendig. Dabei müssten auch hier ergänzende Untersuchungen zu Personenvariablen wie Motivation, aber auch zum Fachwissen und zum themenspezifischen Professionswissen (TPCK) vorgenommen werden, um die Handlungsrelevanz des in Planungen erfassten fachdidaktischen Wissens (PPCK) entsprechend prüfen und einordnen zu können. Längsschnittliche Untersuchungen in dieser Art, wie sie auch von Anthofer (2017) vorgenommen wurden, könnten zudem auch Erkenntnisse zum Professionalisierungsprozess liefern. Wie Zlatkin-Troitschanskaia et al., 2015 erwähnen, sind solche Untersuchungen in der Lehrpersonenausbildung nach wie vor selten. Vor diesem Hintergrund und den in der vorliegenden Studie gewonnenen Erkenntnissen sollen im nächsten Abschnitt mögliche Implikationen für die Lehrpersonenbildung skizziert werden.

II. Implikationen für die Lehrpersonenausbildung

Das professionelle Lehrerhandeln ist in der deutschen erziehungswissenschaftlichen Literatur hinsichtlich des struktur- und kompetenztheoretischen Verständnisses von Professionalität kontrovers diskutiert worden (Kap. I, 2.2.2). Allerdings hat sich der kompetenztheoretische Professionalisierungsansatz bislang insbesondere in der naturwissenschaftsdidaktischen Forschung etabliert (Bölsterli et al., 2010; Cramer, 2012; Reinisch, 2009). Mit Rückblick auf die Ausführungen im letzten Abschnitt zeigt sich, dass auch die Reflexion einen integralen Bestandteil von Professionalität darstellt und eine zentrale Bedeutung in Professionalisierungsprozessen einnimmt (Roters, 2012). Zierer (2015) plädiert deshalb für eine Kombination beider Ansätze, da der kompetenztheoretische Ansatz die Professionalität zu stark auf Wissen und Fähigkeiten reduziert und damit die reflexive Herangehensweise fehlt. Zur Erfassung der oben erwähnten drei Handlungskompetenzen bedarf es daher auch entsprechender Instrumente. In der naturwissenschaftsdidaktischen Forschung hat die Erfassung des Professionswissens mit neu entwickelten, domänenspezifischen Messinstrumenten bereits Tradition und ist auch Gegenstand vieler neuerer Forschungsvorhaben (Krauss et al., 2017). Instrumente, die das *PCK on action* in Unterrichtsplanungen erfassen, sind hingegen rar, genauso wie Instrumente zur Erhebung der fachdidaktischen Qualität von Reflexionen im Sinne von *Reflection on action* (Schön, 1983).

Das im Projekt KUBeX entwickelte und in der vorliegenden Studie eingesetzte Ratingmanual zur Erfassung des fachdidaktischen Wissens zur Planung von Unterricht zum experimentellen Handeln hat sich als faktoriell valide erwiesen. Unter Berücksichtigung der

Umstände, dass es sich dabei um eine erste explorative Validierung handelt und das Instrument weiter geprüft und auch weiterentwickelt werden sollte, scheint dieses Instrument praktikabel für die Lehrpersonenausbildung zu sein. Entsprechend könnte dieses Instrument, das auf der Basis des Modells der «Didaktischen Rekonstruktion» zur Beschreibung und Erfassung eines idealtypischen naturwissenschaftlichen Unterrichts entwickelt worden ist und zwei Phasen der Planung definiert, zum einen als Diagnosetool, zum anderen aber auch als Grundlage für Unterrichtsvorbereitungen eingesetzt werden (Abb. 30). In der Funktion eines Diagnosetools ermöglicht es die Generierung von Informationen zum Professionalisierungsstand der Studierenden in Bezug auf die Vermittlungskompetenz zum kompetenzorientierten Experimentieren. Gleichzeitig können damit auch Rückschlüsse auf die Wirksamkeit der Lehrpersonenbildung und Implikationen für deren Weiterentwicklung gezogen werden.

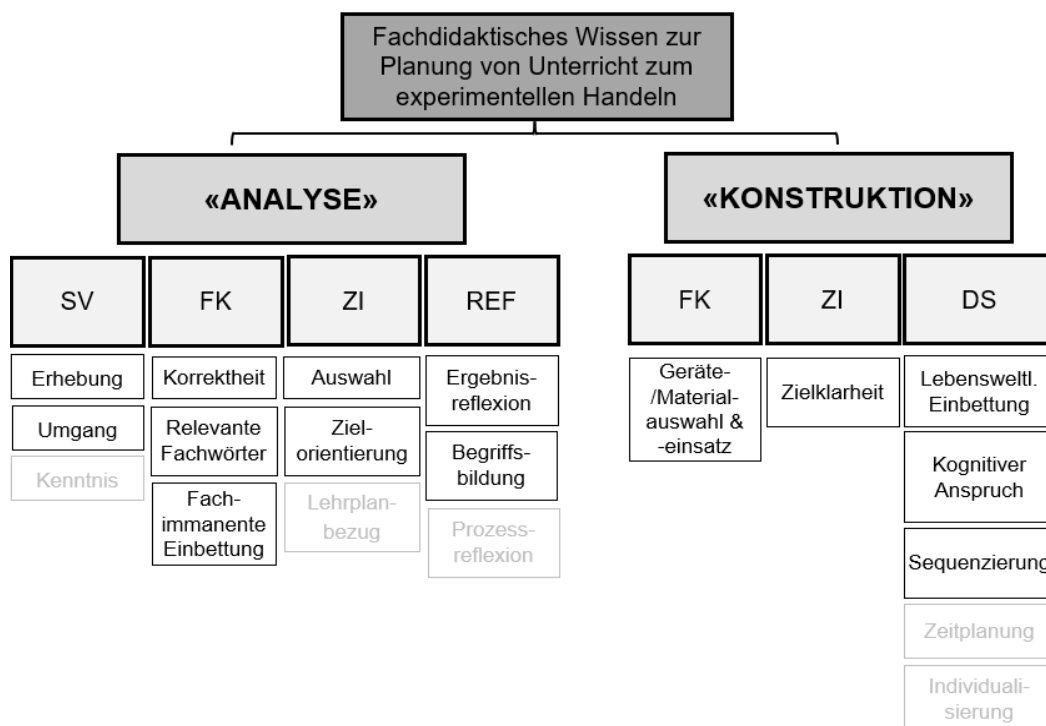


Abbildung 30: Ratingmanual zur Analyse des fachdidaktischen Wissens zur Planung von Unterricht zum experimentellen Handeln unter Berücksichtigung der beiden Dimensionen «Analyse» und «Konstruktion». Hellgrau: Indikatoren, die nicht im Messmodell berücksichtigt werden konnten

Wird das Instrument zur Unterrichtsvorbereitung eingesetzt, so kann es dazu beitragen, dass sich die Studierenden eingehender mit der Planung von Lernprozessen auf der Ebene der Tiefenstruktur des Unterrichts auseinandersetzen. Da den Merkmalen der Tiefenstruktur eine grössere Erklärungsmacht beigemessen wird, was die Lernerfolge der Lernenden

betrifft (Hattie, 2009, Seidel und Shalvelson, 2007), wäre dies anzustreben. Entsprechend müssten die bisherigen Rubriken im Planungsraster konkretisiert und mit bislang noch fehlenden, zentralen Rubriken «Schülervorstellungen» und «Reflexivität» im Sinne einer unterstützenden Massnahme für die Studierenden ergänzt werden. So gesehen, entspräche der Planungsraster auch eher dem Modell, das die Gestaltung eines qualitätsvollen Unterrichts massgeblich unterstützen könnte (cf. Zierer, Wernke & Werner, 2015b). Die Frage, ob mit einem solchen Raster die Qualität der Planungen gesteigert werden könnte, drängt sich hier daher auf. Die Erkenntnis, dass die untersuchte themenspezifische Planungskompetenz der Studierenden zum experimentellen Handeln unterschiedlich ausfiel, wobei tiefere Kompetenzausprägungen vergleichsweise häufiger waren und die Vermittlung des experimentellen Handelns eine Lehrplanvorgabe darstellt, unterstreicht den Handlungsbedarf in der Lehrpersonenausbildung. Im Sinne des dynamischen Modells von Shulman (1987) müsste das fachdidaktische Wissen hierzu in mehreren Durchgängen (Planen-Unterrichten-Reflektieren) aufgebaut werden. Denn wie Mutton et al. (2011) hierzu argumentieren: *„It is through planning that teachers are able to learn about teaching and through teaching that they are able to learn about planning.“*

Anhang

A Teilstudie I

- A-1 Dokumente der Intervention
 - A1.1 Vignette
 - A1.2 Planungsraster
- A-2 Ergänzende Berechnungen zum Forschungsanliegen 1a
 - A2.1 MSA-Werte (*measure of sampling adequacy*)
 - A2.2 Statistische Kennwerte der Indikatoren des Ratingmanuals in der Übersicht (Kappa-gewichtet zum experimentellen Handeln, 2 Rater, 23 Videos)
 - A2.3 Statistische Kennwerte der Kategorien des Ratingmanuals in der Übersicht (ICC_{just} zum experimentellen Handeln, 2 Rater, 23 Videos)

A-1 Dokumente der Intervention



Gesucht:

Stellvertretung im Fach Natur und Technik für eine Klasse der Sekundarstufe I in Zürich

In der Umgebung von Zürich wird aufgrund einer urlaubsbedingten Freistellung einer Lehrperson eine Stellvertretung für eine Doppelstunde in NT gesucht. Die Schule richtet diese Anfrage gezielt an die PHZH. Sie sind darauf aufmerksam geworden und haben sich gemeinsam mit einer/-m Mitstudierenden als Tandem beworben. Die Schule hat sich für Sie entschieden. Gratulation!

Die zu vertretende Lehrperson schickt Ihnen umgehend die unten aufgeführte Liste mit den wesentlichen Angaben zur Vorbereitung.

Bedingungen

Schulhaus:

- gut ausgestattete naturwissenschaftliche Räumlichkeiten
- Experimentiermaterial zum Thema „Visuelle Wahrnehmung“ ist in grosser Zahl vorhanden

Klasse:

- Sekundarstufe I, 8. Klasse, Sek A
- Anzahl Schülerinnen und Schüler: 22, davon 12 Mädchen und 10 Jungen
- SuS mit durchschnittlichem bis gutem Leistungsniveau, sind Unterrichtsmethoden erprobt
- insgesamt gute Arbeitsatmosphäre
- Die SuS bringen erste Experimentiererfahrungen mit. Sie haben bisher vor allem Versuchsaufgaben nach Anleitung gelöst und Resultate in Diagrammen festgehalten.
- Die SuS haben bisher noch keine Experimente eigenständig geplant, durchgeführt und ausgewertet.

Bisher bearbeitete Fachinhalte zum Thema:

- Sinne des Menschen: Haut (Haptische Wahrnehmung), Gehör (Akustische Wahrnehmung), Einführung in den Aufbau des Auges

Planungsauftrag

Thema:	Experimentieren - Lernen am Beispiel Akkommodation oder Adaptation
Zeit:	pro Person eine Doppelstunde (90 Minuten)

Abbildung A1.1: Vignette

Planungsraster für die Unterrichtsskizze

Name		Code
Fachinhalte und Lernziele der Unterrichtssequenz		
Unterrichtsgestaltung zur Unterstützung des beabsichtigten Lernens		
Aktivitäten Lehrperson <i>Formulieren Sie auch Arbeitsaufträge!</i>	Aktivitäten Schüler/innen <i>Lerntätigkeiten</i>	Medien <i>Hilfsmittel aller Art</i>
Weitere Aspekte, die Sie bei der Unterrichtsplanung beachten wollen...		

Abbildung A1.2 Planungsraster

A-2 Ergänzende Berechnungen zum Forschungsanliegen 1a

Tabelle A2.1: MSA-Werte (*measure of sampling adequacy*)

Anti-Image-Korrelation	sui_krs	.458 ^a	-.381	-.021	-.175	-.142	.030	.090	-.136	-.140	.040	.183	.032	.120	.034	.388	-.359
	sui_asu	-.381	.837 ^a	-.350	.022	.228	-.075	-.160	.325	.174	.068	-.361	-.068	-.213	-.080	-.227	.195
	sui_ds	-.021	-.350	.887 ^a	.082	-.417	.235	-.057	-.167	-.138	-.254	.134	.122	.296	-.105	-.002	-.103
	fk_fk	-.175	.022	.082	.962 ^a	-.290	-.055	-.170	.057	.014	-.227	-.046	-.148	.065	-.024	-.024	.107
	fk_fw	-.142	.228	-.417	-.290	.882 ^a	-.310	.053	.105	.252	.193	-.257	.057	-.283	-.203	-.247	.084
	fk_mat	.030	-.075	.235	-.055	-.310	.912 ^a	-.055	-.280	-.280	-.310	.005	-.227	.218	.173	-.033	.045
	ds_fach	.090	-.160	-.057	-.170	.053	-.055	.935 ^a	-.054	-.040	-.037	.030	.208	.113	-.540	-.128	-.163
	ds_leb	-.136	.325	-.167	.057	.105	-.280	-.054	.883 ^a	-.066	.064	-.123	-.036	-.224	-.060	.185	-.154
	ds_kog	-.140	.174	-.138	.014	.252	-.280	-.040	-.066	.943 ^a	-.138	-.267	-.144	.002	-.178	-.004	-.117
	is_seq	.040	.068	-.254	-.227	.193	-.310	-.037	.064	-.138	.917 ^a	-.157	-.092	-.508	-.013	-.191	.201
	is_zeit	.183	-.361	.134	-.046	-.257	.005	.030	-.123	-.267	-.157	.871 ^a	-.049	.048	.185	.211	-.040
	ds_az	.032	-.068	.122	-.148	.057	-.227	.208	-.036	-.144	-.092	-.049	.931 ^a	.155	-.545	-.113	-.063
	is_zk	.120	-.213	.296	.065	-.283	.218	.113	-.224	.002	-.508	.048	.155	.870 ^a	-.306	.172	-.312
	is_zo	.034	-.080	-.105	-.024	-.203	.173	-.540	-.060	-.178	-.013	.185	-.545	-.306	.897 ^a	.055	.121
	ref_res	.388	-.227	-.002	-.024	-.247	-.033	-.128	.185	-.004	-.191	.211	-.113	.172	.055	.874 ^a	-.428
	ref_pro	-.359	.195	-.103	.107	.084	.045	-.163	-.154	-.117	.201	-.040	-.063	-.312	.121	-.428	.768 ^a

Tabelle A2.2: Statistische Kennwerte der Kategorien des Ratingmanuals in der Übersicht (ICC_{just} zum experimentellen Handeln, 2 Rater, 23 Videos)

Kategorien	$ICC_{\text{just}} = .87$ ($F(22) = 7.83, p < .001$)
Schülervorstellungen	$ICC_{\text{just}} = .96$ ($F(22) = 25.67, p < .001$)
Fachliche Klärung	$ICC_{\text{just}} = .87$ ($F(22) = 7.83, p < .001$)
Unterrichtsziele	$ICC_{\text{just}} = .99$ ($F(22) = 274.52, p < .001$)
Didaktische Strukturierung	$ICC_{\text{just}} = .97$ ($F(22) = 39.79, p < .001$)
Reflexivität	$ICC_{\text{just}} = .98$ ($F(22) = 7.83, p < .001$)

Tabelle A2.3: Statistische Kennwerte der Indikatoren des Ratingmanuals in der Übersicht (Kappa-gewichtet zum experimentellen Handeln, 2 Rater, 23 Videos)

Kategorien	Indikatoren	Cohens Kappa	Konfidenzintervall.
Schülervorstellungen	Kenntnis	.65	.01 – 1.00
	Erhebung	.53	.01 – 1.00
	Umgang	.62	.17 – 1.00
Fachliche Klärung	Fachliche Richtigkeit	.63	.42 – .85
	Klärung relevanter Fachwörter	.90	.77 – 1.00
	Material- und Geräteauswahl und -einsatz	.74	.55 – .92
Unterrichtsziele	Zielauswahl	1.00	
	Zielorientierung	.92	.80 – 1.00
	Zielklarheit	.95	.86 – 1.00
Didaktische Strukturierung	Fachimmanente Einbettung	.77	.59 – .95
	Lebensweltliche Einbettung	1.00	
	Kognitiver Anspruch	.86	.71 – 1.00
	Sequenzierung der Lerninhalte	.78	.57 – .96
Reflexivität	Zeitplanung	.75	.46 – 1.00
	Ergebnisreflexion	.78	.47 – 1.00
	Prozessreflexion	1.00	
	Begriffsbildung	.94	.81 – 1.00

B Teilstudie II

- B-1 Ergänzende Berechnungen zum Forschungsanliegen 2a der Teilstudie II
 - B1.1 Mittelwerte der Kategorien im Vergleich (parametrisch, $n = 70$)
 - B1.2 Interkorrelationen der 5 Kategorien (nach Pearson, $n = 70$)
 - B1.3 Deskriptive Statistik der Kategorien der Dimension «Konstruktion» für die Interventionsgruppe (IG, $n = 65$) und Kontrollgruppe (KG, $n = 54$)
 - B1.4 Mittelwerte der Kategorien der Dimension «Konstruktion» im Vergleich ($n = 70$)
- B-2 Ergänzende Berechnungen zum Forschungsanliegen 2b der Teilstudie II
 - B2.1 Korrelationen zwischen dem getesteten fachdidaktischen Wissen (FDW) und dem in Planungsmaterialien erfassten fachdidaktischen Wissen (expH) (parametrisch, $N = 119$)

B-1 Ergänzende Berechnungen zum Forschungsanliegen 2a der Teilstudie II

Tabelle B1.1: Mittelwerte der Kategorien im Vergleich (parametrisch, $n = 70$)

Kategorien	$t(69)$	p	Cohens d
Schülervorstellungen (SV) – Fachliche Klärung (FK)	-11.28	< .001	.81
Schülervorstellungen (SV) – Unterrichtsziele (ZI)	-9.267	< .001	.74
Schülervorstellungen (SV) – Didaktische Strukturierung (DS)	-8.788	< .001	.73
Schülervorstellungen (SV) – Reflexivität (REF)	-1.044	.30	.12
Fachliche Klärung (FK) – Unterrichtsziele (ZI)	1.659	.10	.19
Fachliche Klärung (FK) – Didaktische Strukturierung (DS)	4.338	< .001	.46
Fachliche Klärung (FK) – Reflexivität (REF)	10.742	< .001	.79
Unterrichtsziele (ZI) – Didaktische Strukturierung (DS)	2.889	.005	.33
Unterrichtsziele (ZI) – Reflexivität (REF)	9.251	< .001	.74
Didaktische Strukturierung (DS) – Reflexivität (REF)	7.467	< .001	.67

Tabelle B1.2: Interkorrelationen der 5 Kategorien (nach Pearson, $n = 70$)

Kategorien	SV	FK	ZI	REF	DS
Schülervorstellungen (SV)	–	.61**	.63**	.61**	.59**
Fachliche Klärung (FK)		–	.78**	.66**	.74**
Unterrichtsziele (ZI)			–	.71**	.82**
Reflexivität (REF)				–	.63**
Didaktische Strukturierung (DS)					–

** $p < 0.01$

Tabelle B1.3: Mittelwerte der Kategorien der Dimension «Konstruktion» im Vergleich ($n = 70$)

Dimensionen	<i>z</i>	<i>p</i>
Konstruktion (KON) – Fachliche Klärung (FK)	-1.56	.12
Konstruktion (KON) – Unterrichtsziele (ZI)	-1.33	.18
Konstruktion (KON) – Didaktische Strukturierung (DS)	-1.82	.07

* $p < .05$, ** $p < .01$, *** $p < .001$

Tabelle B1.4: Deskriptive Statistik der Kategorien der Dimension «Konstruktion» für die Interventionsgruppe (IG, $n = 65$) und Kontrollgruppe (KG, $n = 54$)

Dimen- sionen	<i>M</i>	<i>SD</i>	Shapiro-Wilk-Test	<i>M</i>	<i>SD</i>	Shapiro-Wilk-Test
	IG ($n = 65$)			KG ($n = 54$)		
KON-FK	1.65	0.87	$W = 0.37^{***}, p < .001$	1.93	1.00	$W = 0.28^{***}, p < .001$
KON-ZI	1.34	0.67	$W = 0.46^{***}, p < .001$	1.54	0.84	$W = 0.41^{***}, p < .001$
KON-DS	1.48	0.73	$W = 0.31^{**}, p < .001$	1.67	0.73	$W = 0.21^{***}, p < .001$

* $p < .05$, ** $p < .01$, *** $p < .001$

Tabelle B1.5: Mittelwerte der Professionswissenstests (CK, t1 & PCK, t2) für die Interventions- (IG) und Kontrollgruppe (KG) im Vergleich (nichtparametrisch, $N = 119$)

Dimensionen	<i>z</i>	<i>p</i>
Fachwissen (IG, t1) – Fachwissen (KG, t1)	-0.81	.42
Fachdidaktisches Wissen (IG, t2) – Fachdidaktisches Wissen (KG, t2)	-0.39	.69

B-2 Ergänzende Berechnungen zum Forschungsanliegen 2b der Teilstudie II

Tabelle B2.1: Korrelationen zwischen dem getesteten fachdidaktischen Wissen (FDW) und dem in Planungsmaterialien erfassten fachdidaktischen Wissen (expH) (parametrisch, $N = 119$)

Kategorien	<i>M</i>	<i>SD</i>	FDW	expH	ANA	KON
Fachdidaktisches Wissen (FDW)	0.59	0.87	–	.41**	.37**	.43**

** $p < .01$

Abkürzungsverzeichnis

APA	American Psychological Association
ACT	Adaptive Control of Thought
CFI	Comparative Fit Index
CK	Content Knowledge, Fachwissen
COACTIV	Cognitive Activation in the Classroom: The Orchestration of Learning Opportunities for the Enhancement of Insightful Learning in Mathematics (Professionswissen von Lehrkräften, kognitiv aktivierender Mathematikunterricht und die Entwicklung mathematischer Kompetenz)
CxK	Contextual Knowledge, Kontextwissen
D-EDK	Deutscheschweizer Erziehungsdirektorenkonferenz
DFG	Deutsche Forschungsgemeinschaft
EDK	Schweizerische Konferenz der kantonalen Erziehungsdirektoren
EFA	Exploratorische Faktorenanalyse
ERIDOB	European Researchers in Didactics of Biology
ERTE	Educational Reconstruction for Teacher Education
ESEM	Exploratory Structural Equation Modeling
ExMo	Vermittlungs- und Beurteilungskompetenzen zum Experimentieren: Modellierung, Validierung und Messinstrumententwicklung
FDW	Fachdidaktisches Wissen
HarmoS	Harmonisierung der obligatorischen Schule
ICC	Intraklassenkoeffizient
IPN	Leibniz-Institut für die Pädagogik der Naturwissenschaften und Mathematik
KiL	Messung professioneller Kompetenzen in mathematischen und naturwissenschaftlichen Lehramtsstudiengängen
KMK	Kultusministerkonferenz
KMO	Kaiser-Meyer-Olkin-Kriterium

KUBeX	Kollegiales Unterrichtscoaching und Entwicklung experimenteller Kompetenz – Interventionsstudie in der schulpraktischen Ausbildung von Lehrpersonen für Biologie
KUC	Kollegiales Unterrichtscoaching oder Peer Coaching
LMT	Learning Mathematics for Teaching
MPRA	Model of Pedagogical Reasoning and Action
MSA	Measure of Sampling Adequacy
NBPTS	National Board for Professional Teaching Standards
NGSS	Next Generation Science Standards
NRC	National Research Council (Vereinigte Staaten)
PCK	Pedagogical Content Knowledge, fachdidaktisches Wissen
PCK&S	Pedagogical Content Knowledge & Skill
PISA	Programme for International Student Assessment
PK	Pedagogical Knowledge, pädagogisches Wissen
PPCK	Personal Pedagogical Content Knowledge
ProwiN	Professionswissen in den Naturwissenschaften
QuIP	Quality of Instruction in Physics
RMSEA	Root Mean Square Error of Approximation
SBFI	Staatssekretariat für Bildung, Forschung und Innovation
SDDS	Scientific Discovery as Dual Search
TIMSS	Trends in International Mathematics and Science Study
TSPK	Topic-specific professional knowledge
WLS	Weighted Least Squares
WLSMV	Weighted Least Squares Mean and Variance Adjusted
WRMR	Weighted Root Mean Square Residual

Abbildungsverzeichnis

ABBILDUNG 1:	CHRONOLOGISCHER AUFBAU DER ARBEIT MIT DEN FORSCHUNGSANLIEGEN DER TEILSTUDIEN I & II.....	12
ABBILDUNG 2:	HARMoS KOMPETENZMODELL.....	22
ABBILDUNG 3:	KOMPETENZMODELLIERUNG ALS KONTINUUM.....	37
ABBILDUNG 4:	STANDARDS DER LEHRPERSONENBILDUNG NACH TERHART.....	46
ABBILDUNG 5:	MODELL PROFESSIONELLER HANDLUNGSKOMPETENZ – PROFESSIONS- WISSEN (COACTIV).....	49
ABBILDUNG 6:	MODELLTYPEN ZUR ROLLE DES FACHDIDAKTISCHEN WISSENS.....	55
ABBILDUNG 7:	DAS KONSENSMODELL.....	57
ABBILDUNG 8:	SYNOPSIS AUS PROZESSORIENTIERTEN KOMPETENZMODELL UND KONSENSMODELL.....	59
ABBILDUNG 9:	KONZEPTE DES „LEHRERWISSENS“.....	67
ABBILDUNG 10:	FACHSPEZIFISCHES RAHMENMODELL ZUR UNTERRICHTSQUALITÄT.....	74
ABBILDUNG 11:	STRUKTUR DES ERHEBUNGSINSTRUMENTS AUS DEM QUIP-PROJEKT.....	79
ABBILDUNG 12:	TRANSFORMATIONSMODELL ZUM PLANEN VON UNTERRICHT.....	83
ABBILDUNG 13:	MODELL DER «DIDAKTISCHEN REKONSTRUKTION».....	93
ABBILDUNG 14:	ERTE-MODELL.....	95
ABBILDUNG 15:	DESIGN DER TEILSTUDIE I EINGEBETTET IN DAS GESAMTDESIGN DES PROJEKT KUBEX.....	103
ABBILDUNG 16:	KAMERAPOSITIONEN FÜR DIE VIDEOAUFZEICHNUNG DER PLANUNGS- GESPRÄCHE.....	106
ABBILDUNG 17:	ENTWICKLUNGSVERFAHREN DES HOCH INFERENTEN RATINGMANUALS.....	110
ABBILDUNG 18:	CODIER-ANALYSE-KREISLAUF VON VIDEODATEN.....	110
ABBILDUNG 19:	RATINGMANUAL ZUR ANALYSE DES FACHDIDAKTISCHEN WISSENS ZUR PLANUNG VON UNTERRICHT ZUM EXPERIMENTELLEN HANDELN.....	111
ABBILDUNG 20:	RELATIVE HÄUFIGKEITEN DER 4 QUALITÄTSSTUFEN DER 17 INDIKATOREN ZUM FACHDIDAKTISCHEN WISSEN ZUR PLANUNG VON UNTERRICHT ZUM EXPERIMENTELLEN HANDELN (TEILSTICHPROBE N = 70).....	121
ABBILDUNG 21:	1-FAKTORMODELL „FACHDIDAKTISCHES WISSEN ZUR PLANUNG VON UNTERRICHT ZUM EXPERIMENTELLEN HANDELN“ MIT FAKTORLADUNGEN	140
ABBILDUNG 22:	2-FAKTORENMODELL „FACHDIDAKTISCHES WISSENS ZUR PLANUNG VON UNTERRICHT ZUM EXPERIMENTELLEN HANDELN“ MIT FAKTORLADUNGEN	143
ABBILDUNG 23:	2-FAKTORENMODELL MIT STANDARDISIERTEN KOEFFIZIENTEN FÜR DIE FAKTOREN «ANALYSE» UND «KONSTRUKTION» ZUR UNTERSUCHUNG VON UNTERRICHTSPLANUNGEN, -MATERIALIEN UND VIDEOGRAFIERTEN PLANUNGSGESPRÄCHEN VON ANGEHENDEN LEHRERPERSONEN.....	145

ABBILDUNG 24: 2-FAKTORENMODELL (MESSMODELL) MIT BEZUG ZUM PLANUNGSMODELL DER «DIDAKTISCHEN REKONSTRUKTION»	153
ABBILDUNG 25: <i>MODEL OF PEDAGOGICAL REASONING AND ACTION (MPRA)</i>	173
ABBILDUNG 26: DESIGN DER TEILSTUDIE II EINGEBETTET IN DAS GESAMTDESIGN DES PROJEKT KUBEX	182
ABBILDUNG 27: KOMPETENZAUSPRÄGUNGEN ZUR PLANUNG VON UNTERRICHT ZUM EXPERIMENTELLEN HANDELN, GESAMTSTICHPROBE $N = 119$	189
ABBILDUNG 28: KOMPETENZAUSPRÄGUNGEN ZUR PLANUNG VON UNTERRICHT ZUM EXPERIMENTELLEN HANDELN, TEILSTICHPROBE $N = 70$	189
ABBILDUNG 29: MITTELWERTVERGLEICHE DER 5 KATEGORIEN FACHDIDAKTISCHEN WISSENS ZUR PLANUNG VON UNTERRICHT ZUM EXPERIMENTELLEN HANDELN.....	191
ABBILDUNG 30: RATINGMANUAL ZUR ANALYSE DES FACHDIDAKTISCHEN WISSENS ZUR PLANUNG VON UNTERRICHT ZUM EXPERIMENTELLEN HANDELN UNTER BERÜCKSICHTIGUNG DER BEIDEN DIMENSIONEN «ANALYSE» UND «KONSTRUKTION»	219

Tabellenverzeichnis

TABELLE 1:	KOMPETENZBEREICHE DES FACHES BIOLOGIE IN DEUTSCHLAND	21
TABELLE 2:	COACTIV-MODELL – LERNVERSTÄNDNIS UND VERSTÄNDNIS ZUM PROFESSIONSWISSEN	48
TABELLE 3:	MODELLIERUNGEN FACHDIDAKTISCHEN WISSENS BZW. VON PCK.....	62
TABELLE 4:	DETAILLIERTE STICHPROBENBESCHREIBUNG DER VORLIEGENDEN STUDIE (TEILSTUDIE I & II)	104
TABELLE 5:	INDIKATOREN DES RATINGMANUALS MIT AUSGEWÄHLTEN LITERATURBEZÜGEN	113
TABELLE 6:	WERTEBEREICH UND INTERPRETATION VON COHENS KAPPA	117
TABELLE 7:	STATISTISCHE KENNWERTE DER INDIKATOREN DES RATINGMANUALS IN DER ÜBERSICHT	118
TABELLE 8:	ABSOLUTE UND RELATIVE HÄUFIGKEITEN (%) DER VERGEBENEN CODES.....	121
TABELLE 9:	ITEMKENNWERTE DER 17 INDIKATOREN DES RATINGMANUALS ($N = 70$)	123
TABELLE 10:	VALIDITÄTSASPEKTE NACH MESSICK (1995).....	128
TABELLE 11:	MODELL-FIT-STATISTIK 1-FAKTORMODELL	139
TABELLE 12:	MODELL-FIT-STATISTIK 2-FAKTORENMODELL	142
TABELLE 13:	2-FAKTORENMODELL – FAKTORLADUNGEN UND KOMMUNALITÄTEN DER INDIKATOREN.....	144
TABELLE 14:	GOODNESS-OF-FIT-INDIZES DER BEIDEN MODELLE ($N = 70$).....	145
TABELLE 15:	ERREICHTE SUMMENScores ZUM EXPERIMENTELLEN HANDELN	188
TABELLE 16:	DESKRIPTIVE STATISTIK DER 5 KATEGORIEN FACHDIDAKTISCHEN WISSENS ($N = 70$).....	191
TABELLE 17:	MITTELWERTVERGLEICH DER 5 KATEGORIEN ($N = 70$).....	192
TABELLE 18:	INTERKORRELATIONEN DER 5 KATEGORIEN ($N = 70$)	193
TABELLE 19:	DESKRIPTIVE STATISTIK DER ZWEI DIMENSIONEN «ANALYSE» UND «KONSTRUKTION» ($N = 70$)	193
TABELLE 20:	DESKRIPTIVE STATISTIK DER ZWEI DIMENSIONEN FACHDIDAKTISCHEN WISSENS UND DER LATENTEN VARIABLEN EXPH FÜR DIE INTERVENTIONS- GRUPPE (IG, $N = 65$) UND KONTROLLGRUPPE (KG, $N = 54$).....	195
TABELLE 21:	MITTELWERTVERGLEICHE DER DIMENSIONEN (ANA & KON) UND DER LATENTEN VARIABLEN (EXPH) FÜR DIE INTERVENTIONS- (IG) UND KONTROLLGRUPPE (KG).....	195
TABELLE 22:	DESKRIPTIVE STATISTIK DER KATEGORIEN DER DIMENSION «ANALYSE» FÜR DIE INTERVENTIONSGRUPPE (IG, $N = 65$) UND KONTROLLGRUPPE (KG, $N = 54$)	196

TABELLE 23:	MITTELWERTE DER KATEGORIEN DER DIMENSION «ANALYSE» IM VERGLEICH ($N = 70$)	196
TABELLE 24:	KORRELATIONEN ZWISCHEN DEM GETESTETEN FACHDIDAKTISCHEN WISSEN (FDW) UND DEM IN PLANUNGSMATERIALIEN ERFASSTEN FACHDIDAKTISCHEN WISSEN (EXPH) ($N = 119$)	198
TABELLE 25:	KORRELATIONEN ZWISCHEN DEM GETESTETEN FACHWISSEN (FW) UND DEM IN PLANUNGSMATERIALIEN ERFASSTEN FACHDIDAKTISCHEN WISSEN (EXPH) SOWIE DEN BEIDEN DIMENSIONEN «ANALYSE» UND «KONSTRUKTION» ($N = 119$)	199

Literaturverzeichnis

- Abd-El-Khalick, F. & Lederman, N. G. (2000). Improving science teachers' conceptions of nature of science: a critical review of the literature. *International Journal of Science Education*, 22 (7), 665-701.
- Abell, S. K. (2007). Research on science teacher knowledge. In S. K. Abell & N. G. Lederman (Eds.), *Handbook on Research on Science Education* (S. 57-74). New York, London: Routledge.
- Abell, S. K. (2008). Twenty years later: Does pedagogical content knowledge remain a useful idea? *International Journal of Science Education*, 30 (10), 1405-1416.
- Abels, S. (2011). Lehrerinnen und Lehrer als „Reflective Practitioner“ – Die Bedeutsamkeit von Reflexionskompetenz für einen demokratieförderlichen Naturwissenschaftsunterricht. Wiesbaden: Springer Fachmedien.
- Abrahams, I. & Millar, R. (2008). Does practical work really work? A study of the effectiveness of practical work as a teaching and learning method in school science. *International Journal of Science Education*, 30 (14), 1945-1969.
- Abrahams, I. & Reiss, M. J. (2012). Practical work: Its effectiveness in primary and secondary schools in England. *Journal of Research in Science Teaching*, 49 (8), 1035-1055.
- Adamina, M., Wyssen, H., Möller, K., Steffensky, M. & Sunder, C. (2017). Kognitiv anregende und inhaltlich strukturierende Massnahmen der Lernunterstützung im naturwissenschaftlichen Unterricht der Grundschule (in Vorbereitung). *WWU-Portal der Westfälischen Wilhelms-Universität Münster*.
- Alonzo, A. C. & Kim, J. (2016). Declarative and dynamic pedagogical content knowledge as elicited through two video-based interview methods. *Journal of Research in Science Teaching*, 53 (8), 1259-1286.
- American Educational Research Association, American Psychological Association & National Council on Measurement in Education. (2002). *Standards for educational and psychological testing*. Washington DC: American Educational Research Association.
- American Psychological Association (1954). Technical recommendations for psychological tests and diagnostic techniques. *Psychological Bulletin*, 51, 201-238.

- Anderson, J. R. (1976). *Language, memory, and thought*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Anderson, J. R. (2013). *Kognitive Psychologie. Deutsche Ausgabe herausgegeben von Joachim Funke* (7. Auflage.). Berlin, Heidelberg: Springer.
- Anderson, R. D. (2002). Reforming science teaching: What research says about inquiry. *Journal of Science Teacher Education*, 13 (1), 1-12.
- Anthofer, S. (2017). *Förderung des fachspezifischen Professionswissens von Chemielehramtsstudierenden*. Berlin: Logos.
- Arnold, K.-H. & Koch-Priewe, B. (2010). Traditionen der Unterrichtsplanung in Deutschland. *Bildung und Erziehung*, 63 (4), 401-416.
- Arnold, K.-H., Rakhkochkine, A. & Mitter, W. (Hrsg.). (2010). Unterrichtsplanungen im internationalen Vergleich. *Bildung und Erziehung*, 63 (4), 395-400.
- Asparouhov, T. & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 16 (3), 397-438.
- Aufschnaiter, C. v. & Blömeke, S. (2010). Professionelle Kompetenz von (angehenden) Lehrkräften erfassen - Desiderata. *Zeitschrift für Didaktik der Naturwissenschaften*, 16, 361-367.
- Aufschnaiter, C. v. & Rogge, C. (2010). Misconceptions or missing conceptions? *Eurasia Journal of Mathematics, Science & Technology Education*, 6 (1), 3-18.
- Backhaus, K., Erichson, B., Plinke, W. & Weiber, R. (2018). *Multivariate Analysemethoden. Eine anwendungsorientierte Einführung* (15. vollständig überarb. Aufl.). Berlin, Heidelberg: Springer.
- Baer, M., Kocher, M., Wyss, C., Guldimann, T., Larcher, S. & Dörr, G. (2011). Lehrerbildung und Praxiserfahrung im ersten Berufsjahr und ihre Wirkung auf die Unterrichtskompetenzen von Studierenden und jungen Lehrpersonen im Berufseinstieg. *Zeitschrift für Erziehungswissenschaft*, 14, 85-117.
- Ball, D. L. & Bass, H. (2000). Interweaving content and pedagogy in teaching and learning to teach: Knowing and using mathematics. In J. Boaler (Ed.), *Multiple perspectives on the teaching and learning of mathematics* (83-104). Westport, CT: Ablex.
- Ball, D. L., Thames, M. & Phelps, G. (2008). Content knowledge for teaching : What makes it special? *Journal of Teacher Education*, 59, 389-407.

- Bartholomew, H., Osborne, J. & Ratcliffe, M. (2004). Teaching students “ideas-about-science”: Five dimensions of effective practice. *Science Education*, 88 (5), 655-682.
- Baumert, J. & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9 (4), 469-520.
- Baumert, J. & Kunter, M. (2011a). Das Kompetenzmodell von COACTIV. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (29-53). Münster: Waxmann.
- Baumert, J. & Kunter, M. (2011b). Das mathematikspezifische Wissen von Lehrkräften, kognitive Aktivierung im Unterricht und Lernfortschritte von Schülerinnen und Schülern. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (163-192). Münster: Waxmann.
- Baumert, J. & Kunter, M. (2013a). Professionelle Kompetenz von Lehrkräften. In I. Gogolin, H. Kuper, H.-H. Krüger & J. Baumert (Hrsg.), *Stichwort: Zeitschrift für Erziehungswissenschaft* (277-337). Wiesbaden: Springer.
- Baumert, J. & Kunter, M. (2013b). The COACTIV model of teachers’ professional competence. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Eds.), *Cognitive Activation in the Mathematics Classroom and Professional Competence of Teachers. Results from the COACTIV Project* (25-48). Boston: Springer.
- Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A. et al. (2010). Teachers’ mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal*, 47 (1), 133-180.
- Baumert, J., Kunter, M., Blum, W., Klusmann, U., Krauss, S. & Neubrand, M. (2011). Professionelle Kompetenz von Lehrkräften, kognitiv aktivierender Unterricht und die mathematische Kompetenz von Schülerinnen und Schülern (COACTIV) - Ein Forschungsprogramm. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 7-25). Münster: Waxmann.
- Baylor, A. L. (2002). Expanding preservice teachers’ metacognitive awareness of instructional planning through pedagogical agents. *Educational Technology Research and Development*, 50 (2), 5-22.
- Berliner, D. C. (2005). The near impossibility of testing for teacher quality. *Journal of Teacher Education*, 56 (3), 205-213.

- Berry, A., Friedrichsen, P. & Loughran, J. (Eds.). (2015). *Re-examining pedagogical content knowledge in science education*. New York and London: Routledge.
- Besser, M. & Krauss, S. (2009). Zur Professionalität als Expertise. In O. Zlatkin-Troitschanskaia, K. Beck, D. Sembill, R. Nickolaus & R. Mulder (Hrsg.), *Lehrprofessionalität. Bedingungen, Genese, Wirkungen und ihre Messung* (71-82). Weinheim: Beltz.
- Blank, R. (2016). Teacher-students' lesson plan about scientific inquiry in experimental biology pertaining to professional knowledge. Gehalten auf der ERIDOB, Karlstad.
- Blankertz, H. (1969/2000). *Theorien und Modelle der Didaktik* (14. Auflage). Weinheim: Juventa.
- Blömeke, S. (2002). *Universität und Lehrerausbildung*. Bad Heilbrunn: Klinkhardt.
- Blömeke, S. (2006). KMK-Standards für die LehrerInnenbildung in Deutschland – ein Kommentar. *Journal für Lehrerinnen- und Lehrerbildung*, 1 (06), 25-33.
- Blömeke, S., Felbrich, A. & Müller, C. (2008). Theoretischer Rahmen und Untersuchungsdesign. In S. Blömeke, G. Kaiser & R. Lehmann (Hrsg.), *Professionelle Kompetenz angehender Lehrerinnen und Lehrer. Wissen, Überzeugungen und Lerngelegenheiten deutscher Mathematikstudierender und -referendare. Erste Ergebnisse zur Wirksamkeit der Lehrerausbildung* (15-48). Münster: Waxmann.
- Blömeke, S., Gustafsson, J.-E. & Shavelson, R. J. (2015). Beyond Dichotomies. *Zeitschrift für Psychologie*, 223 (1), 3-13.
- Blömeke, S., Kaiser, G. & Lehmann, R. (Hrsg.). (2008). Professionelle Kompetenz angehender Lehrerinnen und Lehrer. Wissen, Überzeugungen und Lerngelegenheiten deutscher Mathematikstudierender und -referendare. Erste Ergebnisse zur Wirksamkeit der Lehrerausbildung. Münster: Waxmann.
- Blömeke, S., Kaiser, G. & Lehmann, R. (Hrsg.). (2010). TEDS-M 2008 – Professionelle Kompetenz und Lerngelegenheiten angehender Mathematiklehrkräfte für die Sekundarstufe I im internationalen Vergleich. Münster: Waxmann.
- Blömeke, S., Kaiser, G., Lehmann, R., König, J., Döhrmann, M. & Buchholtz, C. (2009). TEDS-M: Messung von Lehrerkompetenzen im internationalen Vergleich. In O. Zlatkin-Troitschanskaia, K. Beck, D. Sembill, R. Nickolaus & R. Mulder (Hrsg.), *Lehrprofessionalität. Bedingungen, Genese, Wirkungen und ihre Messung* (181-210). Weinheim: Beltz.

- Blömeke, S. & König, J. (2011). Profile im Professionswissen zur Unterrichtsplanung bei Sekundarstufenlehrkräften. *Jahrbuch für Allgemeine Didaktik*, 11-30.
- Blömeke, S. & Müller, C. (2008). Zum Zusammenhang von Allgemeiner Didaktik und Lehr-Lernforschung im Unterrichtsgeschehen. In M.A. Meyer, M. Prenzel & S. Hellekamps (Hrsg.), *Perspektiven der Didaktik* (Sonderheft 9, 239-258). Wiesbaden: VS.
- Blömeke, S., Paine, L., Houang, R. T., Hsieh, F.-J., Schmidt, W. H., Tatto, M. T. et al. (2008). Future teachers' competence to plan a lesson: First results of a six-country study on the efficiency of teacher education. *ZDM*, 40 (5), 749-762.
- Bodzin, A. M. & Beerer, K. M. (2003). Promoting inquiry-based science instruction: The validation of the science teacher inquiry rubric (STIR). *Journal of Elementary Science Education*, 15 (2), 39-49.
- Bögeholz, S., Joachim, C., Hasse, S. & Hammann, M. (2016). Kompetenzen von (angehenden) Biologielehrkräften zur Beurteilung von Experimentierkompetenzen. *Unterrichtswissenschaft*, 44 (1).
- Bölsterli, K., Brovelli, D., Kauertz, A., Rehm, M., Reinhold, P., Riese, J. et al. (2010). Erfassung der professionellen Kompetenz von Lehramtsstudierenden. *Naturwissenschaftliche Bildung als Beitrag zur Gestaltung partizipativer Demokratie. Gesellschaft für Didaktik der Chemie und Physik, Jahrestagung in Potsdam 2009* (276-278). Berlin: LIT.
- Borko, H., Roberts, S. A. & Shavelson, R. J. J. (2008). Teachers' decision making: From Alan J. Bishop to today. *Critical Issues in Mathematics Education. Major contributions of Alan Bishop* (37-67). New York: Springer.
- Börnin, J. (2012). *Das Experiment als Lerngelegenheit: Vom interkulturellen Vergleich des Physikunterrichts zu Merkmalen seiner Qualität*. Berlin: Logos.
- Borowski, A., Kirschner, S., Liedtke, S. & Fischer, H. E. (2011). Vergleich des Fachwissens von Studierenden, Referendaren und Lehrenden in der Physik. *Physik und Didaktik in Schule und Hochschule*, 1 (10), 1-9.
- Borowski, A., Neuhaus, B. J., Tepner, O., Wirth, J., Fischer, H. E., Leutner, D. et al. (2010). Professionswissen von Lehrkräften in den Naturwissenschaften (ProwiN) - Kurzdarstellung des BMBF-Projekts. *Zeitschrift für Didaktik der Naturwissenschaften*, 16, 341-349.

- Britton, L. R. & Anderson, K. A. (2010). Peer coaching and pre-service teachers: Examining an underutilised concept. *Teaching and Teacher Education*, 26 (2), 306-314.
- Bromme, R. (1981). *Das Denken von Lehrern bei der Unterrichtsvorbereitung. Eine empirische Untersuchung zu kognitiven Prozessen von Mathematiklehrern*. Weinheim und Basel: Beltz.
- Bromme, R. (1995). Was ist „pedagogical content knowledge“? Kritische Anmerkungen zu einem fruchtbaren Forschungsprogramm. In S. Hopmann & K. Riquarts (Hrsg.), *Didaktik und/oder Curriculum* (Zeitschrift für Pädagogik, Beiheft 33, 105-115). Weinheim: Beltz.
- Bromme, R. (1997). Kompetenzen, Funktionen und unterrichtliches Handeln des Lehrers. In F. E. Weinert (Hrsg.), *Psychologie des Unterrichts und der Schule* (S. 177-212). Göttingen: Hogrefe.
- Bromme, R. (2001). Teacher expertise. In N. J. Smelser & P. B. Baltes (Eds.), *International encyclopedia of the social and behavioral sciences* (Vol. 26, 15459-15465). Amsterdam: Elsevier.
- Bromme, R. (2008). Lehrerexpertise. In W. Schneider & M. Hasselhorn (Hrsg.), *Handbuch der Pädagogischen Psychologie* (159-167). Göttingen: Hogrefe.
- Bromme, R. & Rheinberg, F. (2006). Lehrende in Schulen. In A. Krapp & B. Weidenmann (Hrsg.), *Pädagogische Psychologie* (5. Auflage, 296-334). Weinheim: Beltz.
- Bromme, R. & Seeger, F. (1979). *Unterrichtsplanung als Handlungsplanung. Eine psychologische Einführung in die Unterrichtsvorbereitung*. Königstein: Scriptor.
- Brophy, J. E. (2000). „Teaching“. International Academy of Education & International Bureau of Education. Zugriff am 10.3.2019. Verfügbar unter <http://www.ibe.unesco.org/en/document/teaching-educational-practices-1>.
- Brovelli, D., Bölsterli, K., Rehm, M. & Wilhelm, M. (2013). Erfassen professioneller Kompetenzen für den naturwissenschaftlichen Unterricht: Ein Vignettest mit authentisch komplexen Unterrichtssituationen und offenem Antwortformat. *Unterrichtswissenschaft*, 41 (4), 306-329.
- Brovelli, D., Rehm, M. & Wilhelm, M. (2014). Integrierte naturwissenschaftliche Lehrerbildung – Entwicklung professioneller Kompetenz bei Lehramtsstudierenden. In S. Bernholt (Hrsg.), *Naturwissenschaftliche Bildung zwischen Science- und Fachunterricht. Gesellschaft für Didaktik der Chemie und Physik, Jahrestagung in München 2013* (S. 40-54). Kiel: IPN.

- Brown, P., Friedrichsen, P. & Abell, S. (2013). The development of prospective secondary biology teachers PCK. *Journal of Science Teacher Education*, 24 (1), 133-155.
- Brückmann, M. (2009). *Sachstrukturen im Physikunterricht: Ergebnisse einer Videostudie*. Berlin: Logos.
- Brückmann, M. & Bernholt, S. (2013). Videobasierte Erfassung der Komplexitätsentwicklung im Chemie- und Physikunterricht. In U. Riegel & K. Macha (Hrsg.), *Videobasierte Kompetenzforschung in den Fachdidaktiken* (79-96). Münster: Waxmann.
- Brückmann, M. & Duit, R. (2014). Videobasierte Analyse unterrichtlicher Sachstrukturen. In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *Methoden in der naturwissenschaftsdidaktischen Forschung* (189-201). Berlin Heidelberg: Springer.
- Brühwiler, C. (2014). *Adaptive Lehrkompetenz und schulisches Lernen. Effekte handlungssteuernder Kognitionen von Lehrpersonen auf Unterrichtsprozesse und Lernergebnisse der Schülerinnen und Schüler*. Münster: Waxmann.
- Brunner, M., Kunter, M., Krauss, S., Baumert, J., Blum, W., Dubberke, T. et al. (2006). Welche Zusammenhänge bestehen zwischen dem fachspezifischen Professionswissen von Mathematiklehrkräften und ihrer Ausbildung sowie beruflichen Fortbildung? *Zeitschrift für Erziehungswissenschaft*, 9 (4), 521-544.
- Bucat, R. (2004). Pedagogical content knowledge as a way forward: Applied research in chemistry education. *Chemistry Education Research and Practice*, 5 (3), 215-228.
- Bühl, A. (2014). *SPSS 22. Einführung in die moderne Datenanalyse* (14. aktual. Aufl.). Hallbergmoos: Pearson Studium.
- Bühner, M. (2011). *Einführung in die Test- und Fragebogenkonstruktion* (3., aktual. Aufl.). Hallbergmoos: Pearson Studium.
- Bybee, R. W. (1997). *Achieving scientific literacy: From purposes to practices*. Portsmouth, NH: Heinemann.
- Bybee, R. W. (2002). Scientific Literacy – Mythos oder Realität? In W. Gräber, P. Nentwig, T. Koballa & R. Evans (Eds.), *Scientific Literacy. Der Beitrag der Naturwissenschaften zur Allgemeinen Bildung* (21-43). Opladen: Leske + Budrich.
- Capps, D. K. & Crawford, B. A. (2013a). Inquiry-based professional development: What does it take to support teachers in learning about inquiry and nature of science? *International Journal of Science Education*, 35 (12), 1947-1978.

- Capps, D. K. & Crawford, B. A. (2013b). Inquiry-based instruction and teaching about nature of science: Are they happening? *Journal of Science Teacher Education*, 24 (3), 497-526.
- Carlsen, W. S. (1991). Effects of new biology teachers' subject-matter knowledge on curricular planning. *Science Education*, 75 (6), 631-647.
- Carlsen, W. S. (1999). Domains of teacher knowledge. In J. Gess-Newsome & N. G. Lederman (Eds.), *Examining pedagogical content knowledge. The construct and its implication for science education* (133-144). Dordrecht, Boston, London: Kluwer.
- Carnahan, R. S. (1980). *The effects of teacher planning on classroom processes* (Tech. Rep. No. 541). Madison, WI Wisconsin R & D Center for Individualized Schooling.
- Cauet, E. (2016). *Testen wir relevantes Wissen? Zusammenhang zwischen dem Professionswissen von Physiklehrkräften und gutem und erfolgreichem Unterrichten*. Berlin: Logos.
- Causton-Theoharis, J. N., Theoharis, G. T. & Trezek, B. J. (2008). Teaching pre-service teachers to design inclusive instruction: A lesson planning template. *International Journal of Inclusive Education*, 12 (4), 381-399.
- Chabalengula, V. M., Mumba, F. & Mbewe, S. (2012). How pre-service teachers' understand and perform science process skills. *Eurasia Journal of Mathematics, Science & Technology Education*, 8 (3), 167-176.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, Massachusetts: MIT Press.
- Clark, C. M. & Peterson, P. L. (1986). Teachers' thought processes. In M. Wittrock (Ed.), *Handbook of Research on Teaching* (255-296). New York: Macmillan
- Clausen, M. (2002). Unterrichtsqualität - Eine Frage der Perspektive? Münster: Waxmann.
- Cochran, K. F., DeRuiter, J. A. & King, R. A. (1993). Pedagogical content knowing: An integrative model for teacher preparation. *Journal of Teacher Education*, 44 (4), 263-272.
- Cochran-Smith, M. (2001). The outcomes question in teacher education. *Teaching and Teacher Education*, 17 (5), 527-546.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112 (1), 155-159.

- Combe, A. (2006a). Schulentwicklung als Herausforderung für die Lehrerprofessionalität - zur „Individualitätsvergessenheit“ der deutschen Schule. In R. Boenicke, A. Hund, T. Rihm & V. Strittmatter-Haubold (Hrsg.), *Innovativ Schule entwickeln: Kompetenzen, Praxis und Visionen* (47, 37-47). Heidelberg: Mattes.
- Combe, A. (2006b). „Hatten die schon Schuhe?“ Zur Theorie des Erfahrungslernens. *Pädagogik*, 58 (6), 32-36.
- Combe, A. & Paseka, A. (2012). Und sie bewegt sich doch? Gedanken zu Brückenschlägen in der aktuellen Professions- und Kompetenzdebatte. *Zeitschrift für Bildungsforschung*, 2 (2), 91-107.
- Cramer, C. (2012). *Entwicklung von Professionalität in der Lehrerbildung. Empirische Befunde zu Eingangsbedingungen, Prozessmerkmalen und Ausbildungserfahrungen Lehramtstudierender*. Bad Heilbrunn: Julius Klinkhardt.
- Crawford, B. A. (2000). Embracing the essence of inquiry: New roles for science teachers. *Journal of Research in Science Teaching*, 37 (9), 916-937.
- Crawford, B. A. (2007). Learning to teach science as inquiry in the rough and tumble of practice. *Journal of Research in Science Teaching*, 44 (4), 613-642.
- Cronbach, L. J., Rajaratnam, N. & Gleser, G. C. (1963). Theory of generalizability: A liberalization of reliability theory. *British Journal of Statistical Psychology*, 16 (2), 137-163.
- Curran, P. J., West, S. G. & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods*, 1 (1), 16-29.
- Dannemann, S. & Krüger, D. (2010). Evaluation eines Aufgabeninventars zur Ermittlung von Schülervorstellungen zum Sehen [Evaluation of a task inventory for assessing students' concepts of eyesight]. In U. Harms & I. Mackensen-Friedrichs (Hrsg.), *Lehr- und Lernforschung in der Biologiedidaktik* (4, 137-154). Innsbruck: Studienverlag.
- Davies, D. & Rogers, M. (2000). Pre-service primary teachers' planning for science and technology activities: Influences and constraints. *Research in Science and Technological Education*, 18 (2), 215-225.
- Davis, E. A., Petish, D. & Smith, J. (2006). Challenges new science teachers face. *Review of Educational Research*, 76 (4), 607-651.

- DeBoer, G. E. (2000). Scientific literacy: Another look at its historical and contemporary meanings and its relationship to science education reform. *Journal of Research in Science Teaching*, 37 (6), 582-601.
- Deci, E. L. & Ryan, R. M. (2000). The „what“ and „why“ of goal pursuits: Human needs and the selfdetermination of behavior. *Psychological Inquiry*, 11 (4), 227-268.
- D-EDK. (2014). Lehrplan 21: Sprachregionaler Lehrplan für die deutsch- und mehrsprachigen Kantone. Freigegeben zur Einführung an die Kantone am 31. Oktober 2014. Luzern: Geschäftsstelle D-EDK.
- Depaepe, F., Verschaffel, L. & Kelchtermans, G. (2013). Pedagogical content knowledge: A systematic review of the way in which the concept has pervaded mathematics educational research. *Teaching and Teacher Education*, 34, 12-25.
- Derry, S. J., Pea, R. D., Barron, B., Engle, R. A., Erickson, F., Goldman, R. et al. (2010). Conducting video research in the learning sciences: Guidance on selection, analysis, technology, and ethics. *Journal of the Learning Sciences*, 19 (1), 3-53.
- di Fuccia, D.-S., Witteck, T., Markic, S. & Eilks, I. (2012). Trends in practical work in german science education. *Eurasia Journal of Mathematics, Science & Technology Education*, 8 (1), 59-72.
- Dittmer, A. (2010). *Nachdenken über Biologie. Über den Bildungswert der Wissenschaftsphilosophie in der akademischen Biologielehrerbildung*. Wiesbaden: VS.
- Ditton, H. (2009). Unterrichtsqualität. In K.-H. Arnold, U. Sandfuchs & J. Wiechmann (Hrsg.), *Handbuch Unterricht* (2. aktual. Aufl., 177-183). Bad Heilbrunn: Julius Klinkhardt.
- Dorfner, T., Förtsch, C. & Neuhaus, B. J. (2017). Die methodische und inhaltliche Ausrichtung quantitativer Videostudien zur Unterrichtsqualität im mathematisch-naturwissenschaftlichen Unterricht. *Zeitschrift für Didaktik der Naturwissenschaften*, 23, 261-285.
- Döring, N. & Bortz, J. (Hrsg.). (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (5. vollständig überarb., aktual. u. erw. Aufl.). Berlin: Springer.
- Duit, R. (2004). Fachdidaktiken als Forschungsgebiete und als Berufswissenschaften der Lehrkräfte - das Beispiel der Didaktik der Naturwissenschaften. *Beiträge zur Lehrerinnen- und Lehrerbildung*, 22 (1), 20-28.

- Duit, R. (2007). Science education research internationally: Conceptions, research methods, domains of research. *Eurasia Journal of Mathematics, Science & Technology Education*, 3 (1), 3-15.
- Duit, R., Gropengiesser, H., Kattmann, U., Komorek, M. & Parchmann, I. (2012). The model of educational reconstruction - a framework for improving teaching and learning science. *Science Education Research and Practice in Europe* (5, 13-37). Rotterdam: Sense.
- Duit, R., Treagust, D. F. & Widodo, A. (2008). Teaching science for conceptual change - theory and practice. In S. Vosniadou (Ed.), *International Handbook of Research on Conceptual Change*, 629-646. New York: Routledge.
- Duschl, R. (2008). Science Education in three-part harmony: Balancing conceptual, epistemic, and social learning goals. *Review of Research in Education*, 32 (1), 268-291.
- Duschl, R. A. (2003). Assessment of inquiry. In J.M. Atkin & J.E. Coffey (Hrsg.), *Everyday Assessment in the Science Classroom* (S. 41-59). Arlington, VA: NSTA Press.
- Duschl, R. A. & Wright, E. (1989). A case study of high school teachers' decision making models for planning and teaching science. *Journal of Research in Science Teaching*, 26 (6), 467-501.
- EDK (2011). Grundkompetenzen für die Naturwissenschaften: Nationale Bildungsstandards. Freigegeben von der EDK-Plenarversammlung am 16. Juni, 2011. Bern. http://edudoc.ch/record/96787/files/grundkomp_nawi_d.pdf. Zugriffen: 17.8.2016.
- Einsiedler, W. (2002). Das Konzept Unterrichtsqualität. *Unterrichtswissenschaft*, 3, 194-196.
- Emden, M. (2011). *Prozessorientierte Leistungsmessung des naturwissenschaftlich-experimentellen Arbeitens. Eine vergleichende Studie zu Diagnoseinstrumenten zu Beginn der Sekundarstufe I*. Berlin: Logos.
- Ergönenc, J., Neumann, K. & Fischer, H. E. (2014). The impact of pedagogical content knowledge on cognitive activation and student learning. In H.E. Fischer, P. Labudde, K. Neumann & J. Viiri (Hrsg.), *Quality of Instruction in Physics – comparing Finland, Germany and Switzerland* (145-160). Münster: Waxmann.
- Fazio, X., Melville, W. & Bartley, A. (2010). The problematic nature of the practicum: A key determinant of pre-service teachers' emerging inquiry-based science practices. *Journal of Science Teacher Education*, 21 (6), 665-681.

- Fend, H. (1998). *Qualität im Bildungswesen. Schulforschung zu Systembedingungen, Schulprofilen und Lehrerleistung*. Weinheim: Juventa.
- Fend, H. (2002). Mikro- und Makrofaktoren eines Angebot-Nutzungsmodells von Schulleistungen. *Zeitschrift für Pädagogische Psychologie*, 16 (3/4), 141-149.
- Fend, H. (2008). Dimensionen von Qualität im Bildungswesen. Von Produktindikatoren zu Prozessindikatoren am Beispiel der Schule. In E. Klieme & R. Tippelt (Hrsg.), *Qualitätssicherung im Bildungswesen, Zeitschrift für Pädagogik, Beiheft 53* (190-209). Weinheim: Beltz.
- Fend, H. (2011). Die Wirksamkeit der neuen Steuerung – theoretische und methodische Probleme ihrer Evaluation. *Zeitschrift für Bildungsforschung*, 1, 5-24.
- Fensham, P. J. (1986). Science for all. *Educational Leadership*, 44 (4), 18-23.
- Fensham, P. J. (2001). Science content as problematic – Issues for research. In H. Behrendt, H. Dahncke, R. Duit, W. Gräber, M. Komorek & A. Kross (Eds.), *Research in Science Education – Past, Present, and Future* (27-41). Dordrecht: Kluwer.
- Fenstermacher, G. D. (1994). The knower and the known: The nature of knowledge in research on teaching. *Review of Research in Education* (20, 3-56). Washington DC: American Educational Research Association.
- Fernandez, C. (2014). Knowledge base for teaching and pedagogical content knowledge (PCK): some useful models and implications for teachers' training. *Problems of education in the 21st century*, 60, 79-100.
- Field, A. (2013). *Discovering Statistics using IBM SPSS Statistics*. London: SAGE.
- Finney, S. J. & DiStefano, C. (2006). Nonnormal and categorical data in structural equation modeling. In G. R. Hancock & R. O. Mueller (Eds.), *Structural Equation Modeling: A second course* (269-314). Charlotte, NC: Information Age.
- Fischer, H. E. (1998). Scientific Literacy und Physiklernen. *Zeitschrift für Didaktik der Naturwissenschaften*, (2), 41-52.
- Fischer, H. E., Klemm, K., Leutner, D., Sumfleth, E., Tiemann, R. & Wirth, J. (2003). Naturwissenschaftsdidaktische Lehr-Lernforschung: Defizite und Desiderata. *Zeitschrift für Didaktik der Naturwissenschaften*, 9, 179-208.
- Fischler, H. (2007). Fachdidaktik und Unterrichtsqualität im Bereich Naturwissenschaften. In K.-H. Arnold (Hrsg.), *Unterrichtsqualität und Fachdidaktik* (235-260). Bad Heilbrunn: Klinkhardt.

- Fischler, H. (2008). Physikdidaktisches Wissen und Handlungskompetenz. *Zeitschrift für Didaktik der Naturwissenschaften*, 14, 27-49.
- Fischler, H. (2011). Didaktik - An appropriate framework for the professional work of science teachers? D. Corrigan, J. Dillon & R. Gunstone (Eds.), *The Professional Knowledge Base of Science Teaching* (31-50). Berlin: Springer.
- Förtsch, C., Werner, S., Kotzebue, L. von & Neuhaus, B. J. (2016). Effects of biology teachers' professional knowledge and cognitive activation on students' achievement. *International Journal of Science Education*, 38 (17), 2642-2666.
- Frank, C., Bernholt, S. & Parchmann, I. (2016). Modellierung des Zusammenhangs allgemeiner und beruflicher Kompetenzen für die Domäne Chemie. *Zeitschrift für Didaktik der Naturwissenschaften*, 22 (1), 43-60.
- Frey, A. (2006). Methoden und Instrumente zur Diagnose beruflicher Kompetenzen von Lehrkräften. Eine erste Standortbestimmung zu bereits publizierten Instrumenten. In C. Allemann-Ghionda & E. Terhart (Hrsg.), *Kompetenzen und Kompetenzentwicklung von Lehrerinnen und Lehrern: Ausbildung und Beruf*. Weinheim: Beltz.
- Frey, A. (2014). Kompetenzmodelle und Standards in der Lehrerbildung und im Lehrerberuf. In E. Terhart, H. Bennewitz & M. Rothland (Hrsg.), *Handbuch der Forschung zum Lehrerberuf* (712-744). Münster: Waxmann.
- Frudden, S. J. (1984). Lesson plans can make a difference in evaluating teachers. *Education*, 104 (4), 351-53.
- Furtak, E. M., Seidel, T., Iverson, H. & Briggs, D. C. (2012). Experimental and quasi-experimental studies of inquiry-based science teaching: A meta-analysis. *Review of Educational Research*, 82 (3), 300-329.
- Futter, K. (2017). *Lernwirksame Unterrichtsbesprechungen im Praktikum*. Heilbrunn: Julius Klinkhardt.
- Gardner, A. L. & Gess-Newsome, J. (2011). A PCK rubric to measure teachers' knowledge of inquiry-based instruction using three data sources. *A paper presented at the Annual Meeting of the National Association for Research in Science Teaching*. Orlando, FL.
- Gassmann, C. (2013). *Erlebte Aufgabenschwierigkeit bei der Unterrichtsplanung*. Wiesbaden: Springer.
- Gebauer, M. M. (2013). *Determinanten der Selbstwirksamkeitsüberzeugung von Lehrenden. Schulischer Berufsalltag an Gymnasien und Hauptschulen*. Wiesbaden: Springer.

- Geiser, C. (2011). *Datenanalyse mit MPlus. Eine anwendungsorientierte Einführung* (2., durchgesehene Aufl.). Wiesbaden: VS.
- George, D. & Mallery, P. (2010). *SPSS for Windows step by step: A simple guide and reference, 17.0 update* (10th ed.). Boston: Pearson.
- Gess-Newsome, J. & Lederman, N. G. (1995). Biology teachers' perceptions of subject matter structure and its relationship to classroom practice. *Journal of Research in Science Teaching*, 32 (3), 301-325.
- Gess-Newsome, J. (1999). Pedagogical content knowledge: An introduction and orientation. The nature and history of pedagogical knowledge. In J. Gess-Newsome & N. G. Lederman (Eds.), *Examining pedagogical content knowledge. The construct and its implications for science education* (3-17). Dordrecht: Kluwer.
- Gess-Newsome, J. (2015). A model of teacher professional knowledge and skill including PCK. In A. Berry, P. Friedrichsen & J. Loughran (Eds.), *Re-examining pedagogical content knowledge in science education* (28-42). New York and London: Routledge.
- Gess-Newsome, J., Taylor, J. A., Carlson, J., Gardner, A. L., Wilson, C. D. & Stuhlsatz, M. A. M. (2017). Teacher pedagogical content knowledge, practice, and student achievement. *International Journal of Science Education*, 1-20.
- Goldston, M. J., Dantzler, J., Day, J. & Webb, B. (2013). A psychometric approach to the development of a 5E lesson plan scoring instrument for inquiry-based teaching. *Journal of Science Teacher*, 24, 527-551.
- Gomez-Zwiep, S. (2008). Elementary teachers' understanding of students' science misconceptions: Implications for practice and teacher education. *Journal of Science Teacher Education*, 19 (5), 437-454.
- Gott, R. & Duggan, S. (2002). Problems with the assessment of performance in practical science: Which way now? *Cambridge Journal of Education*, 32 (2), 183-201.
- Gräber, W. & Nentwig, P. (2002). Scientific Literacy – Naturwissenschaftliche Grundbildung in der Diskussion. In W. Gräber, P. Nentwig, T. Koballa & R. Evans (Hrsg.), *Scientific Literacy. Der Beitrag der Naturwissenschaften zur Allgemeinen Bildung* (7-20). Opladen: Leske + Budrich.
- Gramzow, Y. (2015). *Fachdidaktisches Wissen von Lehramtsstudierenden im Fach Physik. Modellierung und Testkonstruktion*. Berlin: Logos.

- Gramzow, Y., Riese, J. & Reinhold, P. (2013). Modellierung fachdidaktischen Wissens angehender Physiklehrkräfte. *Zeitschrift für Didaktik der Naturwissenschaften*, 19, 7-30.
- Gräsel, C. & Trempler, K. (Hrsg.). (2017). *Entwicklung von Professionalität pädagogischen Personals. Interdisziplinäre Betrachtungen, Befunde und Perspektiven*. Wiesbaden: Springer.
- Greiten, S. (2015). Modellierung von Kompetenzen zur Unterrichtsplanung mit dem Schwerpunkt der individuellen Förderung. In U. Riegel, S. Schubert, G. Siebert-Ott & K. Macha (Hrsg.), *Kompetenzmodellierung und Kompetenzmessung in den Fachdidaktiken*, 243-255. Münster: Waxmann.
- Gropengiesser, H. (1997). Didaktische Rekonstruktion des «Sehens». Wissenschaftliche Theorien und die Sicht der Schüler in der Perspektive der Vermittlung (durchges.u. überarb. als Bd 1 der BzDR (2001, 2007)). Oldenburg: ZpB.
- Grossman, P. L. (1990). The making of a teacher: Teacher knowledge and teacher education. *Journal of Teacher Education*, 42 (5), 379-382.
- Grossman, P. L., Wilson, S. M. & Shulman, L. S. (1989). Teachers of substance: Subject matter knowledge for teaching. In M.C. Reynolds (Ed.), *Knowledge base for the beginning teacher* (23-36). New York: Pergamon.
- Grossschedl, J. & Harms, U. (2013). Assessing conceptual knowledge using similarity judgments. *Studies in Educational Evaluation*, 39 (2), 71-81.
- Grossschedl, J., Harms, U., Kleickmann, T. & Glowinski, I. (2015). Preservice biology teachers' professional knowledge: Structure and learning opportunities. *Journal of Science Teacher Education*, 26 (3), 291-318.
- Grossschedl, J., Neubrand, C., Kirchner, A., Oppermann, L., Basel, N. & Gantner, S. (2015). Entwicklung und Validierung eines Testinstruments zur Erfassung des evolutionsbezogenen Professionswissens von Lehramtsstudierenden (ProWiE). *Zeitschrift für Didaktik der Naturwissenschaften*, 21 (1), 173-185.
- Gruehn, S. (2000). *Unterricht und schulisches Lernen : Schüler als Quellen der Unterrichtsbeschreibung*. Münster: Waxmann.
- Gut, C., Hild, P., Metzger, S. & Tardent, J. (2014). Projekt ExKoNawi: Modell für hands-on Assessments experimenteller Kompetenzen. In S. Bernholt (Hrsg.), *Naturwissenschaftliche Bildung zwischen Science- und Fachunterricht. Gesellschaft für Didaktik der Chemie und Physik, Jahrestagung in München 2013* (171-173). Kiel: IPN.

- Gut, C., Metzger, S., Hild, P. & Tardent, J. (2014). Problemtypenbasierte Modellierung und Messung experimenteller Kompetenzen von 12- bis 15-jährigen Jugendlichen. *Didaktik der Physik, Frühjahrstagung-Frankfurt 2014*, 1-9.
- Gut-Glanzmann, C. (2012). *Modellierung und Messung experimenteller Kompetenz: Analyse eines large-scale Experimentiertests*. Berlin: Logos.
- Gyllenpalm, J. & Wickman, P.-O. (2011a). The uses of the term hypothesis and the inquiry emphasis conflation in science teacher education. *International Journal of Science Education*, 33 (14), 1993-2015.
- Gyllenpalm, J. & Wickman, P.-O. (2011b). "Experiments" and the inquiry emphasis conflation in science teacher education. *Science Education*, 95 (5), 908-926.
- Gyllenpalm, J., Wickman, P.-O. & Holmgren, S.-O. (2010). Teachers' language on scientific inquiry: Methods of teaching or methods of inquiry? *International Journal of Science Education*, 32 (9), 1151-1172.
- Haas, A. (1998). *Unterrichtsplanung im Alltag: eine empirische Untersuchung zum Planungs Handeln von Hauptschul-, Realschul- und Gymnasiallehrern*. Regensburg: Roderer.
- Haas, A. (2005). Unterrichtsplanung im Alltag von Lehrerinnen und Lehrern. In A. A. Huber (Hrsg.), *Vom Wissen zum Handeln. Ansätze zur Überwindung der Theorie-Praxis-Kluft in Schule und Erwachsenenbildung* (5-19). Tübingen: Huber.
- Habig, S., van Vorst, H. & Sumfleth, E. (2018). Merkmale kontextualisierter Lernaufgaben und ihre Wirkung auf das situationale Interesse und die Lernleistung von Schülerinnen und Schülern. *Zeitschrift für Didaktik der Naturwissenschaften*, 24 (1), 99-114.
- Hacker, R. & Sova, B. (1998). Initial teacher education: A study of the efficacy of computer mediated courseware delivery in a partnership context. *British Journal of Educational Technology*, 29 (4), 333-341.
- Hallitzky, M., Marchand, S. & Seibert, N. (2013). Ausgewählte didaktische Modelle und ihre Bedeutung für eine theoriegeleitete Erforschung und Praxis des Unterrichts. In L. Haag, S. Rahm, H. J. Apel & W. Sacher (Hrsg.), *Studienbuch Schulpädagogik* (5. Aufl.). Bad Heilbrunn: Klinkhardt.
- Hammann, M. (2004). Kompetenzentwicklungsmodelle. Merkmale und ihre Bedeutung - dargestellt anhand von Kompetenzen beim Experimentieren. *Der mathematische und naturwissenschaftliche Unterricht*, 57 (4), 196-203.

- Hammann, M., Jördens, J., Krüger, D., Parchmann, I. & Schecker, H. (2014). Offene Aufgaben codieren. *Methoden der naturwissenschaftsdidaktischen Forschung*. Berlin: Springer.
- Hammann, M., Phan, T. H. & Bayrhuber, H. (2008). Experimentieren als Problemlösen: Lässt sich das SDDS- Modell nutzen, um unterschiedliche Dimensionen beim Experimentieren zu messen? *Zeitschrift für Erziehungswissenschaften, Sonderheft Kompetenzdiagnostik* (33-49). Wiesbaden: VS.
- Hammann, M., Phan, T. T., Ehmer, M. & Grimm, T. (2008). Assessing pupils' skills in experimentation. *Journal of Biological Education*, 42 (2), 66-72.
- Hardy, I., Jonen, A., Möller, K. & Stern, E. (2006). Effects of instructional support within constructivist learning environments for elementary school students' understanding of "floating and sinking". *Journal of Educational Psychology*, 98 (2), 307-328.
- HarmoS: Konsortium Naturwissenschaften. (2008). *HarmoS Naturwissenschaften⁺: Kompetenzmodell und Vorschläge für Bildungsstandards (Wissenschaftlicher Schlussbericht)*. Wissenschaftliches Konsortium HarmoS Naturwissenschaften.
- Harms, U. & Riese, J. (2018). Professionelle Kompetenz und Professionswissen. In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *Theorien in der naturwissenschaftsdidaktischen Forschung* (283-298). Berlin, Heidelberg: Springer.
- Hartig, D. J. & Klieme, P. D. E. (2006). Kompetenz und Kompetenzdiagnostik. In P. D. K. Schweizer (Hrsg.), *Leistung und Leistungsdiagnostik* (127-143). Heidelberg: Springer.
- Hartig, J. (2008). Kompetenzen als Ergebnis von Bildungsprozessen. *Bundesministerium für Bildung und Forschung* (15-25). Bonn: BMBF.
- Hartig, J. & Frey, A. (2013). Sind Modelle der Item-Response-Theorie (IRT) das „Mittel der Wahl“ für die Modellierung von Kompetenzen? *Zeitschrift für Erziehungswissenschaft*, 16 (1), 47-51.
- Hartig, J., Frey, A. & Jude, N. (2012). Validität. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (2., aktualisierte u. überarb. Aufl., 143-172). Heidelberg: Springer.
- Hashweh, M. Z. (2005). Teacher pedagogical constructions: a reconfiguration of pedagogical content knowledge. *Teachers and Teaching*, 11 (3), 273-292.

- Hasse, S., Joachim, C., Bögeholz, S. & Hammann, M. (2014). Assessing teaching and assessment competences of biology teacher trainees: Lessons from item development. *International Journal of Education in Mathematics, Science and Technology*, 2 (3), 191-205.
- Hattie, J. (2009). *Visible learning - A synthesis of meta-analyses related to achievement*. London: Routledge.
- Hattie, J. & Yates, G. C. R. (2015). *Lernen sichtbar machen. Überarbeitete deutschsprachige Ausgabe von „Visible Learning“ besorgt von Wolfgang Beywl und Klaus Zierer*. Baltmannsweiler: Schneider Hohengehren.
- Hatton, N. & Smith, D. (1995). Reflection in teacher education: Towards definition and implementation. *Teaching and Teacher Education*, 11 (1), 33-49.
- Herbart, J. F. (1806). *Allgemeine Pädagogik aus dem Zweck der Erziehung abgeleitet*. Göttingen: Röwer.
- Heimann, P. (1962). Didaktik als Theorie und Lehre. *Die Deutsche Schule*, 54, 407-427.
- Heimann, P., Otto, G. & Schulz, W. (Hrsg.). (1965). *Unterricht – Analyse und Planung*. Hannover: Schroedel.
- Helmke, A. (2012). *Unterrichtsqualität und Lehrerprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts* (4. überarb. Auflage). Seelze: Klett-Kallmeyer.
- Helsper, W. (2007). Eine Antwort auf Jürgen Baumerts und Mareike Kunters Kritik am strukturtheoretischen Professionsansatz. *Zeitschrift für Erziehungswissenschaft*, 10 (4), 567-579.
- Helsper, W. (2011). Lehrerprofessionalität - der strukturtheoretische Professionsansatz zum Lehrerberuf. In E. Terhart, H. Bennewitz & M. Rothland (Hrsg.), *Handbuch der Forschung zum Lehrerberuf* (149-170). Münster: Waxmann.
- Helsper, W. & Hummrich, M. (2008). Arbeitsbündnis, Schulkultur und Milieu – Reflexionen zu Grundlagen schulischer Bildungsprozesse. In G. Breidenstein & F. Schütze (Hrsg.), *Paradoxien in der Reform der Schule: Ergebnisse qualitativer Sozialforschung* (43-72). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Helsper, W. & Kolbe, F.-U. (2002). Bachelor/Master in der Lehrerbildung — Potential für Innovation oder ihre Verhinderung? *Zeitschrift für Erziehungswissenschaft*, 5 (3), 384-400.
- Henke, K. (2010). A model for memory systems based on processing modes rather than consciousness. *Nature Review Neuroscience*, 11 (7), 523-532.

- Henze, I. & van Driel, J. H. (2015). Toward a more comprehensive way to capture PCK in its complexity. In A. Berry, P. Friedrichsen & J. Loughran (Eds.), *Re-examining pedagogical content knowledge in science education* (120-134). New York: Routledge.
- Herking, T.-V. (2015). *Der kompetente und professionelle Unternehmensberater*. Wiesbaden: Springer.
- Herweg, C. (2008). *Zielorientierung im deutschen und schweizerischen Physikunterricht – eine Videostudie*. Elektronische Dissertation. Kiel: Universität. Zugriff am 3.1.2019. Verfügbar unter: https://macau.uni-kiel.de/servlets/MCRFileNodeServlet/dissertation_derivate_00002440/dissertation_constanze_herweg.pdf
- Heymann, H. W. (2007). Unterricht vorbereiten und planen. *Pädagogik*, 59 (10), 6-9.
- Hill, H. C., Rowan, B. & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal*, 42 (2), 371-406.
- Hodson, D. (1996). Laboratory work as scientific method: three decades of confusion and distortion. *Journal of Curriculum Studies*, 28 (2), 115-135.
- Hodson, D. (2014). Learning science, learning about science, doing science: Different goals demand different learning methods. *International Journal of Science Education*, 36 (15), 2534-2553.
- Hofstein, A. & Lunetta, V. N. (2004). The laboratory in science education: Foundations for the twenty-first century. *Science Education*, 88 (1), 28-54.
- Holbrook, J. & Rannikmae, M. (2009). The meaning of scientific literacy. *International Journal of Environmental and Science Education*, 4 (3), 275-288.
- Höttecke, D. & Riess, F. (2015). Naturwissenschaftliches Experimentieren im Lichte der jüngeren Wissenschaftsforschung – Auf der Suche nach einem authentischen Experimentbegriff der Fachdidaktik. *Zeitschrift für Didaktik der Naturwissenschaften*, 21, 127-139.
- Hurd, P. D. (1958). Science Literacy: Its meaning for American schools. *Educational Leadership*, 16 (1), 13-16.
- IBM Corp. (2016). *IBM SPSS Statistics for Windows, Version 24.0*. Armonk, NY: IBM Corp.

- Irion, T. (2010). Hypercoding in der empirischen Lehr-Lern-Forschung. Möglichkeiten der synchronen Analyse multicodaler Datensegmente zur Rekonstruktion subjektiver Perspektiven in Videostudien. M. Corsten, M. Krug, C. Moritz (Hrsg.), *Videographie praktizieren. Herangehensweisen, Möglichkeiten und Grenzen* (139-161). Wiesbaden: VS.
- Jacobs, C. L., Martin, S. N. & Otieno, T. C. (2008). A science lesson plan analysis instrument for formative and summative program evaluation of a teacher education program. *Science Education*, 92 (6), 1096-1126.
- Jacobs, J., Garnier, H., Gallimore, R., Hollingsworth, H., Givvin, K. B., Trigwell, K. et al. (2003). *TIMSS 1999 Video Study Technical Report: Volume 1, Mathematics Study*. Washington DC: U.S. Department of Education: National Center for Education Statistics.
- Jacobs, J. K., Kawanaka, T. & Stigler, J. W. (1999). Integrating qualitative and quantitative approaches to the analysis of video data on classroom teaching. *International Journal of Educational Research*, 31, 717-724.
- Jäncke, L. (2017). *Lehrbuch kognitive Neurowissenschaften* (2. überarb. Aufl.). Bern: Hogrefe.
- Järvinen, H.-S. (2014). *Lehrerprofessionalisierung durch schulische Netzwerke? Theoretische Ansatzpunkte und empirische Hinweise*. Dortmund: Technische Universität Dortmund.
- Jatzwauk, P., Rumann, S. & Sandmann, A. (2008). Der Einfluss des Aufgabeneinsatzes im Biologieunterricht auf die Lernleistung der Schüler – Ergebnisse einer Videostudie. *Zeitschrift für Didaktik der Naturwissenschaften*, 14, 263-283.
- Jerrim, J., Oliver, M. & Sims, S. (2019). The relationship between inquiry-based teaching and students' achievement. New evidence from a longitudinal PISA study in England. *Learning and Instruction*, 61, 35-44.
- John, P. D. (2006). Lesson planning and the student teacher: Re-thinking the dominant model. *Journal of Curriculum Studies*, 38 (4), 483-498.
- de Jong, T. & Ferguson-Hessler, M. G. M. (1996). Types and qualities of knowledge. *Educational Psychologist*, 31 (2), 105-113.

- Jüttner, M., Boone, W. J., Park S. & Neuhaus, B. J. (2013). Development and use of a test instrument to measure biology teachers' content knowledge (CK) and pedagogical content knowledge (PCK). *Educational Assessment, Evaluation and Accountability*, 25 (1), 45-67.
- Jüttner, M. & Neuhaus, B. J. (2012). Development of items for a pedagogical content knowledge test based on empirical analysis of pupils' errors. *International Journal of Science Education*, 34 (7), 1125-1143.
- Jüttner, M. & Neuhaus, B. J. (2013). Das Professionswissen von Biologielehrkräften. Ein Vergleich zwischen Biologielehrkräften, Biologen und Pädagogen. *Zeitschrift für Didaktik der Naturwissenschaften*, 19, 31-49.
- Kalyuga, S., Rikers, R. & Paas, F. (2012). Educational implications of expertise reversal effects in learning and performance of complex cognitive and sensorimotor skills. *Educational Psychology Review*, 24 (2), 313-337.
- Kamm, E. & Bieri, C. (2008). Forschung in der Lehrerinnen- und Lehrerbildung – professionstheoretische Bezugspunkte zur Konzeption der Master-Thesis in der Ausbildung von Lehrpersonen der Sekundarstufe I. *Beiträge zur Lehrerbildung*, 26 (1), 85-100.
- Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational Measurement* (17-64). ACE: Praeger.
- Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50 (1), 1-73.
- Kang, N.-H. (2008). Learning to teach science: Personal epistemologies, teaching goals, and practices of teaching. *Teaching and Teacher Education*, 24 (2), 478-498.
- Kanning, U. P. (2004). *Standards der Personaldiagnostik. Personalauswahl professionell gestalten*. Göttingen: Hogrefe.
- Kansanen, P. (2009). Subject-matter didactics as a central knowledge base for teachers, or should it be called pedagogical content knowledge? *Pedagogy, Culture & Society*, 17 (1), 29-39.
- Kattmann, U. (2003). Vom Blatt zum Planeten – Scientific Literacy und kumulatives Lernen im Biologieunterricht und darüber hinaus. In B. Moschner, H. Kiper & U. Kattmann (Hrsg.), *PISA 2000 als Herausforderung* (115-137). Baltmannsweiler: Schneider Hohengehren.

- Kattmann, U., Duit, R., Gropengiesser, H. & Komorek, M. (1997). Das Modell der Didaktischen Rekonstruktion - Ein Rahmen für naturwissenschaftsdidaktische Forschung und Entwicklung. *Zeitschrift für Didaktik der Naturwissenschaften*, 3, 3-18.
- Kaufhold, M. (2006). *Kompetenz und Kompetenzerfassung. Analyse und Beurteilung von Verfahren der Kompetenzerfassung*. Wiesbaden: VS.
- Kempa, R. (1986). *Assessment in Science*. Cambridge: Cambridge University Press.
- Kind, V. (2009). Pedagogical content knowledge in science education: Perspectives and potential for progress. *Studies in Science Education*, 45 (2), 169-204.
- Kiper, H. (2011). Unterrichtsplanung auf der Grundlage einer Integrativen Didaktik. *Jahrbuch für Allgemeine Didaktik*, 125-142.
- Kiper, H. & Mischke, W. (2009). *Unterrichtsplanung. Bachelor, Master*. Weinheim: Beltz.
- Kirschner, S., Borowski, A., Fischer, H. E., Gess-Newsome, J. & Aufschnaiter, C. v. (2016). Developing and evaluating a paper-and-pencil test to assess components of physics teachers' pedagogical content knowledge. *International Journal of Science Education*, 38 (8), 1343-1372.
- Kirschner, S., Sczudlek, M., Tepner, O., Borowski, A., Fischer, H. E., Lenske, G. et al. (2017). Professionswissen in den Naturwissenschaften (ProwiN). In C. Gräsel & K. Templer (Hrsg.), *Entwicklung von Professionalität pädagogischen Personals Interdisziplinäre Betrachtungen, Befunde und Perspektiven* (113). Wiesbaden: Springer.
- Kizil, A. & Kattmann, U. (2013). Ein neues Design fürs Experimentieren. Eine empirische Untersuchung. In M. Komorek & S. Prediger (Hrsg.), *Der lange Weg zum Unterrichtsdidaktischen Design. Zur Begründung und Umsetzung fachdidaktischer Forschungs- und Entwicklungsprogramme* (187-201). Münster: Waxmann.
- Klafki, W. (1958). Didaktische Analyse als Kern der Unterrichtsvorbereitung. *Die Deutsche Schule*, 50 (10), 450-471.
- Klafki, W. (1980). Zur Unterrichtsplanung im Sinne kritisch-konstruktiver Didaktik. In B. Adl-Amini & R. Künzli (Hrsg.), *Didaktische Modelle der Unterrichtsplanung* (11-48). Weinheim: Juventa.
- Klafki, W. (1999). Die bildungstheoretische Didaktik im Rahmen kritisch-konstruktiver Erziehungswissenschaft. Oder: Zur Neufassung der Didaktischen Analyse. In H. Gudjons & R. Winkel (Hrsg.), *Didaktische Theorien* (13-34). Hamburg: Westermann.
- Klahr, D. (2000). *Exploring Science: The cognition and development of discovery processes*. Massachusetts: MIT.

- Klieme, E. (2006). Empirische Unterrichtsforschung: aktuelle Entwicklungen, theoretische Grundlagen und fachspezifische Befunde. *Zeitschrift für Pädagogik*, 52 (6), 765-773.
- Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber, H., Prenzel, M. et al. (2003). *Zur Entwicklung nationaler Bildungsstandards. Eine Expertise*. Berlin: Bundesministerium für Bildung und Forschung (BMBF). Zugriff am 31.12.2016. Verfügbar unter: <http://www.dipf.de/de/pdf-dokumente/projekte-materialien/zur-entwicklung-nationaler-bildungsstandards>
- Klieme, E. & Hartig, J. (2007). Kompetenzkonzepte in den Sozialwissenschaften und im erziehungswissenschaftlichen Diskurs. In M. Prenzel, I. Gogolin & H.-H. Krüger (Hrsg.), *Kompetenzdiagnostik* (11-29). Wiesbaden: VS.
- Klieme, E. & Leutner, D. (2006). Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen. Beschreibung eines neu eingerichteten Schwerpunktprogramms der DFG. *Zeitschrift für Pädagogik*, 52 (6), 876-903.
- Klieme, E. & Rakoczy, K. (2008). Empirische Unterrichtsforschung und Fachdidaktik. Outcome-orientierte Messung und Prozessqualität des Unterrichts. *Zeitschrift für Pädagogik*, 54 (2), 222-237.
- Kline, R. B. (2016). *Principles and Practice of Structural Equation Modeling* (4th ed.). New York: The Guilford Press.
- Klusmann, U. & Kunter, M. (2008). *Die Balance finden: Berufliches Engagement und die Fähigkeit zur Distanzierung bei Lehrkräften*. Jahrbuch. Berlin.
- Klusmann, U., Kunter, M., Trautwein, U. & Baumert, J. (2006). Lehrerbelastrung und Unterrichtsqualität aus der Perspektive von Lehrenden und Lernenden. *Zeitschrift für Pädagogische Psychologie*, 20 (3), 161-173.
- KMK. (2004). Standards für die Lehrerbildung: Bildungswissenschaften.
- KMK. (2005). Bildungsstandards im Fach Biologie für den Mittleren Schulabschluss (Jahrgangsstufe 10). Zugriff am 17.8.2016. Verfügbar unter: http://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2004/2004_12_16-Bildungsstandards-Biologie.pdf.
- Koeppen, K., Hartig, J., Klieme, E. & Leutner, D. (2008). Current issues in competence modeling and assessment. *Zeitschrift für Psychologie*, (216), 61-73.

- Komorek, M. & Kattmann, U. (2008). The model of educational reconstruction. In S. Mikelskis-Seifert, U. Ringelband & M. Brückmann (Hrsg.), *Four Decades of Research in Science Education - from Curriculum Development to Quality Improvement* (S. 162-179). Münster: Waxmann.
- König, J. (2010). Lehrerprofessionalität - Konzepte und Ergebnisse der internationalen und deutschen Forschung am Beispiel fachübergreifender, pädagogischer Kompetenzen. In J. König & B. Hofmann (Hrsg.), *Professionalität von Lehrkräften - Was sollen Lehrkräfte im Lese- und Schreibunterricht wissen und können?* (40-105). Berlin: DGLS.
- König, J. (2015). Kontextualisierte Erfassung von Lehrerkompetenzen. *Zeitschrift für Pädagogik*, 61 (3), 305-309.
- König, J., Buchholtz, C. & Dohmen, D. (2015). Analyse von schriftlichen Unterrichtsplanungen: Empirische Befunde zur didaktischen Adaptivität als Aspekt der Planungskompetenz angehender Lehrkräfte. *Zeitschrift für Erziehungswissenschaft*, 18, 375-404.
- Korneck, F., Kohlenberger, M., Oettinghaus, L., Kunter, M. & Lamprecht, J. (2013). Lehrerüberzeugungen und Unterrichtshandeln im Fach Physik. *PhyDid B, Didaktik der Physik - Beiträge zur DPG-Frühjahrstagung, 2013*.
- Kotzebue, L. v. & Nerdel, C. (2012). Professionswissen von Biologielehrkräften zum Umgang mit Diagrammen. *Zeitschrift für Didaktik der Naturwissenschaften*, 18.
- Kotzebue, L. v. & Nerdel, C. (2015). Modellierung und Analyse des Professionswissens zur Diagrammkompetenz bei angehenden Biologielehrkräften. *Zeitschrift für Erziehungswissenschaft*, 18 (4), 687-712.
- Kotzebue, L. v. & Neuhaus, B. J. (2016). Was macht einen guten Unterricht und einen guten Lehrer aus? Trends der Unterrichtsqualitäts- und Lehrerprofessionalitätsforschung. In A. Sandmann & P. Schmiemann (Hrsg.), *Biologiedidaktische Forschung: Schwerpunkte und Forschungsstände* (117-142). Berlin: Logos.
- Krammer, K., Schnetzler, C. L., Pauli, C., Reusser, K., Ratzka, N., Lipowsky, F. et al. (2010). Unterrichtsvideos in der Lehrerfortbildung: Überblick über Konzeption und Ergebnisse einer einjährigen netzgestützten Fortbildungsveranstaltung. *Lehrerinnen und Lehrer lernen: Konzepte und Befunde zur Lehrerfortbildung* (227-243). Münster: Waxmann.

- Krapp, A. (2000). Interest and human development during adolescence. An educational-psychological approach. In J. Heckhausen (Hrsg.), *Motivational psychology of human development* (109-128). London: Elsevier.
- Krapp, A. & Prenzel, M. (2011). Research on interest in science: Theories, methods, and findings. *International Journal of Science Education*, 33 (1), 27-50.
- Krauss, S. (2011). Das Experten-Paradigma in der Forschung zum Lehrerberuf. In E. Terhart, H. Bennewitz & M. Rothland (Hrsg.), *Handbuch der Forschung zum Lehrerberuf* (171-191). Münster: Waxmann.
- Krauss, S., Blum, W., Brunner, M., Neubrand, M., Baumert, J. & Kunter, M. (2011). Konzeptualisierung und Testkonstruktion zum fachbezogenen Professionswissen von Mathematiklehrkräften. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (135-161). Münster: Waxmann.
- Krauss, S. & Bruckmaier, G. (2014). Das Experten-Paradigma in der Forschung zum Lehrerberuf. In E. Terhart, H. Bennewitz & M. Rothland (Hrsg.), *Handbuch der Forschung zum Lehrerberuf* (2. überarb. u. erweit. Aufl., 241-261). Münster: Waxmann.
- Krauss, S., Brunner, M., Kunter, M., Baumert, J., Neubrand, M. & Jordan, A. (2008). Pedagogical content knowledge and content knowledge of secondary mathematics teachers. *Journal of Educational Psychology*, 100 (3), 716-725.
- Krauss, S., Kunter, M., Brunner, M., Baumert, J., Blum, W., Neubrand, M. et al. (2004). COACTIV: Professionswissen von Lehrkräften, kognitiv aktivierender Mathematikunterricht und die Entwicklung von mathematischer Kompetenz. In J. Doll & M. Prenzel (Hrsg.), *Bildungsqualität von Schule: Lehrerprofessionalisierung, Unterrichtsentwicklung und Schulförderung als Strategien der Qualitätsverbesserung* (31-53). Münster: Waxmann.
- Krauss, S., Lindl, A., Schilcher, A. & Tepner, O. (2017). Das Forschungsprojekt FALKO – ein einleitender Überblick. In S. Krauss, A. Lindl, A. Schilcher, M. Fricke, A. Göhring, B. Hofmann et al. (Hrsg.), *FALKO: Fachspezifische Lehrkompetenzen. Konzeption von Professionswissenstests in den Fächern Deutsch, Englisch, Latein, Physik, Musik, Evangelische Religion und Pädagogik*. Münster: Waxmann.
- Krauss, S., Neubrand, M., Blum, W., Baumert, J., Brunner, M., Kunter, M. et al. (2008). Die Untersuchung des professionellen Wissens deutscher Mathematik-Lehrerinnen und -Lehrer im Rahmen der COACTIV-Studie. *Journal für Mathematik-Didaktik*, 29 (3/4), 223-258.

- Kreis, A. & Schnebel, S. (Hrsg.). (2017). *Peer Coaching in der praxissituierten Ausbildung von Lehrpersonen* (Lehrerbildung auf dem Prüfstand) (Band Sonderheft). Landau: Empirische Pädagogik.
- Kreis, A. & Staub, F. (2007). Förderung der Betreuungsarbeit in der berufspraktischen Ausbildung von Lehrpersonen durch fachspezifisches Unterrichtscoaching. In D. Flagemeyer & M. Rotermund (Hrsg.), *Mehr Praxis in der Lehrerbildung - aber wie? Möglichkeiten zur Verbesserung und Evaluation der Lehrerbildung* (95-114). Leipzig.
- Kreis, A. & Staub, F. (2009). Kollegiales Unterrichtscoaching als Methode kooperativen Lernens von Lehrpersonen. In K. Maag Merki (Hrsg.), *Kooperation und Netzwerkbildung. Strategien zur Qualitätsentwicklung in Einzelschulen* (26-39). Seelze: Klett-Kallmeyer.
- Kreis, A. & Staub, F. (2012). Lernen zukünftiger Lehrpersonen im Kontext von Unterrichtsbesprechungen im Praktikum – multiple Indikatoren für ein schwer zu fassendes Phänomen. In M. Gläser-Zikuda, T. Seidel, C. Rohlf, A. Gröschner & S. Ziegelbauer (Hrsg.), *Mixed-Methods in der Bildungsforschung* (209-226). Münster: Waxmann.
- Kreis, A. & Staub, F. C. (2011). Fachspezifisches Unterrichtscoaching im Praktikum. *Zeitschrift für Erziehungswissenschaft*, 14 (1), 61.
- Kröger, J., Euler, M., Neumann, K., Härtig, H. & Petersen, S. (2012). Messung Professioneller Kompetenz im Fach Physik. In S. Bernholt (Hrsg.), *Konzepte fachdidaktischer Strukturierung für den Unterricht. Gesellschaft für Didaktik der Chemie und Physik. Jahrestagung in Oldenburg 2011* (616-618). Berlin: LIT.
- Kröger, J., Neumann, K. & Petersen, S. (2013). Messung professioneller Kompetenz im Fach Physik. *Inquiry-based learning - Forschendes Lernen* (533-535). Kiel: IPN.
- Kron, F. W., Jürgens, E. & Standop, J. (2014). *Grundwissen Didaktik* (6. überarb. Aufl.). München, Basel: Ernst Reinhardt.
- Kulgemeyer, C. (2013). Gelingensbedingungen physikalischer Erklärungen. *PhyDid B - Didaktik der Physik - Beiträge zur DPG-Frühjahrstagung 2013*.
- Kulgemeyer, C., Borowski, A., Fischer, H., Gramzow, Y., Reinhold, P., Riese, J. et al. (2012). ProfiLe-P – Professionswissen in der Lehramtsausbildung Physik - Vorstellung eines Forschungsverbundes. *Didaktik der Physik, Frühjahrstagung-Mainz*, 1-6.

- Künsting, J., Billich, M. & Lipowsky, F. (2009). Der Einfluss von Lehrerkompetenzen und Lehrerhandeln auf den Schulerfolg von Lernenden. In O. Zlatkin-Troitschanskaia, K. Beck, D. Sembill, N. Reinhold & R. Mulder (Hrsg.), *Lehrerprofessionalität. Bedingungen, Genese, Wirkungen und ihre Messung* (655-667). Weinheim: Beltz.
- Kunter, M. (2011). Motivation als Teil der professionellen Kompetenz - Forschungsbefunde zum Enthusiasmus von Lehrkräften. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (259-275). Münster: Waxmann.
- Kunter, M. & Baumert, J. (2006). Who is the expert? Construct and criteria validity of student and teacher ratings of instruction. *Learning Environments Research*, 9 (3), 231-251.
- Kunter, M. & Ewald, S. (2016). Bedingungen und Effekte von Unterricht: Aktuelle Forschungsperspektiven aus der pädagogischen Psychologie. In N. McElvany, W. Bos, H. G. Holtappels, M. M. Gebauer & F. Schwabe (Hrsg.), *Bedingungen und Effekte guten Unterrichts* (Bd. 1, 9-20). Münster, New York: Waxmann.
- Kunter, M., Klusmann, U. & Baumert, J. (2009). Professionelle Kompetenz von Mathematiklehrkräften: Das COACTIV-Modell. In O. Zlatkin-Troitschanskaia, K. Beck, R. Nickolaus & R. Mulder (Hrsg.), *Lehrprofessionalität. Bedingungen, Genese, Wirkungen und ihre Messung* (153-165). Weinheim: Beltz.
- Kunter, M., Klusmann, U., Baumert, J., Richter, D., Voss, T. & Hachfeld, A. (2013). Professional competence of teachers: Effects on instructional quality and student development. *Journal of Educational Psychology*, 105 (3), 805-820.
- Kunter, M. & Trautwein, U. (2013). *Psychologie des Unterrichts*. Stuttgart: UTB.
- Kunter, M., Tsai, Y.-M., Klusmann, U., Brunner, M., Krauss, S. & Baumert, J. (2008). Students' and mathematics teachers' perceptions of teacher enthusiasm and instruction. *Learning and Instruction*, 18 (5), 468-482.
- Kunter, M. & Voss, T. (2011). Das Modell der Unterrichtsqualität in COACTIV: Eine multikriteriale Analyse. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (85-113). Münster: Waxmann.
- Kunz, H. (2011). *Professionswissen von Lehrkräften der Naturwissenschaften im Kompetenzbereich Erkenntnisgewinnung*. Kassel: Universität Kassel.

- Kyriakides, L., Christoforou, C. & Charalambous, C. Y. (2013). What matters for student learning outcomes: A meta-analysis of studies exploring factors of effective teaching. *Teaching and Teacher Education*, 36, 143-152.
- Labudde, P. & Adamina, M. (2008). HarmoS Naturwissenschaften: Impulse für den naturwissenschaftlichen Unterricht von morgen. *Beiträge zur Lehrerbildung*, 26 (3), 351-360.
- Labudde, P. & Möller, K. (2012). Stichwort: Naturwissenschaftlicher Unterricht. *Zeitschrift für Erziehungswissenschaft*, (15), 11-36.
- Lagler, E. & Wilhelm, M. (2013). Zusammenhang von Schülerleistung und Fachausbildung der Lehrkräfte in den Naturwissenschaften – eine Pilotstudie zur Situation in der Schweiz. *Chimica & Didacticae*, 105 (38), 47-70.
- Lange, K., Kleickmann, T., Tröbst, S. & Möller, K. (2012). Fachdidaktisches Wissen von Lehrkräften und multiple Ziele im naturwissenschaftlichen Sachunterricht. *Zeitschrift für Erziehungswissenschaft*, 15, 55-75.
- Lazonder, A. W. & Harmsen, R. (2016). Meta-analysis of inquiry-based learning: Effects of guidance. *Review of Educational Research*, 86 (3), 681-718.
- Lederman, N. G. & Gess-Newsome, J. (1992). Do subject matter knowledge, pedagogical knowledge, and pedagogical content knowledge constitute the ideal gas law of science teaching? *Journal of Science Teacher Education*, 3 (1), 16-20.
- Lederman, N. G., Lederman, J. S. & Antink, A. (2013). Nature of science and scientific inquiry as contexts for the learning of science and achievement of scientific literacy. *International Journal of Education in Mathematics, Science and Technology*, 1 (3), 138-147.
- Lee, E. & Luft, J. A. (2008). Experienced secondary science teachers' representation of pedagogical content knowledge. *International Journal of Science Education*, 30 (10), 1343-1363.
- Leuders, T. (2014). Modellierungen mathematischer Kompetenzen – Kriterien für eine Validitätsprüfung aus fachdidaktischer Sicht. *Journal für Mathematik-Didaktik*, 35 (1), 7-48.
- Leuders, T., Nückles, M., Mikelskis-Seifert, S. & Philipp, K. (Hrsg.). (2019). *Pädagogische Professionalität in Mathematik und Naturwissenschaften. Beiträge zu Struktur, Genese und Wirkungen von Kompetenzen von Lehrkräften*. Wiesbaden: Springer.

- Leutner, D., Hartig, J. & Jude, N. (2008). Measuring competencies: Introduction to concepts and questions of assessment in education. In J. Hartig, E. Klieme & D. Leutner (Eds.), *Assessment of Competencies in Educational Contexts* (177-192). Göttingen: Hogrefe.
- Lipowsky, F. (2006). Auf den Lehrer kommt es an. Empirische Evidenzen für Zusammenhänge zwischen Lehrerkompetenzen, Lehrerhandeln und dem Lernen der Schüler, *Zeitschrift für Pädagogik*, 51. Beiheft, 47-70).
- Lipowsky, F. (2009). Unterrichtsentwicklung durch Fort- und Weiterbildungsmaßnahmen für Lehrpersonen. *Beiträge zur Lehrerinnen- und Lehrerbildung*, 27 (3), 346-360.
- Lipowsky, F. (2015). Unterricht. In E. Wild & J. Möller (Hrsg.), *Pädagogische Psychologie* (69-105). Berlin: Springer.
- Lipowsky, F., Rakoczy, K., Pauli, C., Drollinger-Vetter, B., Klieme, E. & Reusser, K. (2009). Quality of geometry instruction and its short-term impact on students' understanding of the Pythagorean Theorem. *Learning and Instruction*, 19 (6), 527-537.
- Lunetta, V. N., Hofstein, A. & Clough M. P. (2005). Learning and teaching in the school science laboratory: An analysis of research, theory and practice. In S. K. Abell & N. G. Lederman (Eds.), *Handbook of Research on Science Education*, (393-441). Mahwah, NJ.: Lawrence Erlbaum.
- Maag Merki, K. (2004). Überfachliche Kompetenzen als Ziele beruflicher Bildung im betrieblichen Alltag. *Zeitschrift für Pädagogik*, 50 (2), 202-222.
- Magnusson, S., Krajcik, J. & Borko, H. (1999). Nature, sources, and development of pedagogical content knowledge for science teaching. In J. Gess-Newsome & N. G. Lederman (Hrsg.), *Examining Pedagogical Content Knowledge* (95-132). Dordrecht: Kluwer.
- Marsh, H. W., Morin, A. J. S., Parker, P. D. & Kaur, G. (2014). Exploratory structural equation modeling: An integration of the best features of exploratory and confirmatory factor analysis. *Annu. Rev. Clin. Psychol.*, 10, 85-110.
- Mayer, J. (2007). Erkenntnisgewinnung als wissenschaftliches Problemlösen. In D. Krüger & H. Vogt (Hrsg.), *Theorien in der biologiedidaktischen Forschung: Ein Handbuch für Lehramtsstudenten und Doktoranden* (177-186). Berlin: Springer.
- Mayer, J., Grube, C. & Möller, A. (2008). Kompetenzmodell naturwissenschaftlicher Erkenntnisgewinnung. In U. Harms & A. Sandmann (Hrsg.), *Lehr- und Lernforschung in der Biologiedidaktik* (Bd 3, 63-79). Innsbruck: Studienverlag.

- Mayer, J. & Wellnitz, N. (2014). Die Entwicklung von Kompetenzstrukturmodellen. In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *Methoden in der naturwissenschaftsdidaktischen Forschung* (19-29). Berlin, Heidelberg: Springer.
- Mayr, W. (2011). Der Persönlichkeitsansatz in der Lehrerforschung. Konzepte, Befunde und Folgerungen. In E. Terhart, H. Bennewith & M. Rothland (Hrsg.), *Handbuch der Forschung zum Lehrerberuf* (125-148). Münster: Waxmann.
- Mayring, P. (2015). *Qualitative Inhaltsanalyse. Grundlagen und Techniken* (12. überarb. Aufl.). Weinheim: Beltz.
- Melville, W., Fazio, X., Bartley, A. & Jones, D. (2008). Experience and reflection: Pre-service science teachers' capacity for teaching inquiry. *Journal of Science Teacher Education*, 19 (5), 477-494.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, (50), 741-749.
- Messmer, R. (2015). Stimulated recall as a focused approach to action and thought processes of teachers - Stimulated Recall als fokussierter Zugang zu Handlungs- und Denkprozessen von Lehrpersonen. *Forum Qualitative Sozialforschung*, 16 (1), 160-169.
- Meyer, M. A. (2009). Unterrichtsplanung aus der Perspektive der Bildungsgangforschung. M.A. Meyer, M. Prenzel, S. Hellekamps (Hrsg.), *Perspektiven der Didaktik*, Sonderheft 9 (117-137). Wiesbaden: VS.
- Millar, R. (1996). Designing a curriculum for public understanding of science. *Education in Science*, 8-10.
- Millar, R. (2004). The role of practical work in the teaching and learning of science. *High School Science Laboratories: Role and Vision*. Washington, DC: National Academy of Sciences.
- Millar, R. (2006). Twenty first century science: Insights from the design and implementation of a scientific literacy approach in school science. *International Journal of Science Education*, 28 (13), 1499-1521.
- Moosbrugger, H. & Kelava, A. (2012). Qualitätsanforderungen an einen psychologischen Test (Testgütekriterien). In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (2., aktual. u. überarb. Aufl., 8-26). Heidelberg: Springer.

- Mugaloglu, E. & Saribas, D. (2010). Pre-service science teachers' competence to design an inquiry based lab lesson. *Procedia - Social and Behavioral Sciences*, 2 (2), 4255-4259.
- Muthén, L. K. & Muthén, B. O. (1998, 2017). *Mplus user's guide (7th ed.)*. Los Angeles, CA: Muthén & Muthén.
- Mutton, T., Hagger, H. & Burn, K. (2011). Learning to plan, planning to learn: the developing expertise of beginning teachers. *Teachers and Teaching*, 17 (4).
- National Board for Professional Teaching Standards [NBPTS]. (2002). What teachers should know and be able to do. Arlington, VA. Zugriff am 1.3. 2019. Verfügbar unter: http://www.nbpts.org/sites/default/files/what_teachers_should_know.pdf
- National Research Council (NRC). (1996). *National science educational standards*. Washington DC: The National Academies Press.
- National Research Council (NRC). (2012). *A framework for K-12 science education. Practices, crosscutting concepts, and core ideas*. Washington, DC: The National Academies Press.
- Neuhaus, B. (2007). Unterrichtsqualität als Forschungsfeld für empirische biologiedidaktische Studien. In D. Krüger & H. Fischler (Hrsg.), *Theorien in der biologiedidaktischen Forschung* (43-254). Berlin: Springer.
- Neumann, K. (2013). Mit welchem Auflösungsgrad können Kompetenzen modelliert werden? In welcher Beziehung stehen Modelle zueinander, die Kompetenz in einer Domäne mit unterschiedlichem Auflösungsgrad beschreiben? *Zeitschrift für Erziehungswissenschaft*, 16 (1), 35-39.
- Neuweg, G. H. (2005). Emergenzbedingungen pädagogischer Könnerschaft. In H. Heid & C. Harteis (Hrsg.), *Verwertbarkeit: Ein Qualitätskriterium (erziehungs-)wissenschaftlichen Wissens?* (205-228). Wiesbaden: VS.
- Neuweg, G. H. (2014). Das Wissen der Wissensvermittler - Problemstellungen, Befunde und Perspektiven der Forschung zum Lehrerwissen. In E. Terhart, H. Bennewitz & M. Rothland (Hrsg.), *Handbuch der Forschung zum Lehrerberuf* (2. überarb. u. erweit. Aufl.). Münster: Waxmann.
- NGSS Lead States. (2013). *Next generation science standards: For states, by states*. Washington D.C: The National Academies Press.
- Nijveldt, M. (2007). *Validity in teacher assessment. An exploration of the judgement processes of assessors*. Leiden University.

- Nilsson, P. & Vikström, A. (2015). Making PCK explicit—Capturing science teachers' pedagogical content knowledge (PCK) in the science classroom. *International Journal of Science Education*, 37 (17), 2836-2857.
- Nittel, D. (2004). Die „Veralltäglichung“ pädagogischen Wissens - im Horizont von Profession, Professionalisierung und Professionalität. *Zeitschrift für Pädagogik*, 50 (3), 342-357.
- Nitz, S., Nerdel, C. & Prechtel, H. (2012). Entwicklung eines Erhebungsinstruments zur Erfassung der Verwendung von Fachsprache im Biologieunterricht. *Zeitschrift für Didaktik der Naturwissenschaften*, 18, 117-139.
- Nowak, K. H., Nehring, A., Tiemann, R. & Upmeyer zu Belzen, A. (2013). Assessing students' abilities in processes of scientific inquiry in biology using a paper-and-pencil test. *Journal of Biological Education*, 47 (3), 182-188.
- Oberauer, K. (1993). Prozedurales und deklaratives Wissen und das Paradigma der Informationsverarbeitung. *Sprache und Kognition*, 12, 30-43.
- OECD. (2007). *PISA 2006: Schulleistungen im internationalen Vergleich: Naturwissenschaftliche Kompetenzen für die Welt von Morgen*. Paris: OECD.
- Oelkers, J. & Oser, F. (2000). *Die Wirksamkeit der Lehrerbildungssysteme in der Schweiz. Umsetzungsbericht*. Aarau: SKBF.
- Oelkers, J. & Reusser, K. (2008). *Expertise: Qualität entwickeln – Standards sichern – mit Differenz umgehen*. Berlin: BMBF. Zugriff am 1.3. 2019. Verfügbar unter <http://www.skbf-csre.ch/information/nfp33/ub.oelkers.pdf>
- Oevermann, U. (1996). Theoretische Skizze einer revidierten Theorie professionalisierten Handelns. Untersuchungen zum Typus pädagogischen Handelns. In A. Combe & W. Helsper (Hrsg.), *Pädagogische Professionalität. Untersuchungen zum Typus pädagogischen Handelns* (71-182). Frankfurt a. M.: Suhrkamp.
- Ohle, A. (2010). *Primary school teachers' content knowledge in physics and its impact on teaching and students' achievement. Studien zum Physik- und Chemielernen*. Berlin: Logos.
- Ohle, A., Fischer, H. E. & Kauertz, A. (2011). Der Einfluss des physikalischen Fachwissens von Primarstufenlehrkräften auf Unterrichtsgestaltung und Schülerleistung. *Zeitschrift für Didaktik der Naturwissenschaften*, 17, 357-389.
- Olszewski, J. (2010). *The impact of physics teachers' pedagogical content knowledge on teacher actions and student outcomes*. Berlin: Logos.

- Oser, F. (1997). Standards in der Lehrerbildung. Teil 1: Berufliche Kompetenzen, die hohen Qualitätsmerkmalen entsprechen. *Beiträge zur Lehrerinnen- und Lehrerbildung*, 15 (1), 26-37.
- Oser, F. (2001). Modelle der Wirksamkeit in der Lehrer- und Lehrerinnenbildung. In F. Oser & J. Oelkers (Hrsg.), *Die Wirksamkeit der Lehrerbildungssysteme. Von der Allrounderbildung zur Ausbildung professioneller Standards* (67-96). Chur, Zürich: Rüegger.
- Oser, F. (2009). Moral jenseits von organisierter Erlaubtheit – Zur inneren und äusseren Effizienz eines professionellen Ethos. In O. Zlatkin-Troitschanskaia, K. Beck, D. Sembill, R. Nickolaus & R. Mulder (Hrsg.), *Lehrerprofessionalität. Bedingungen, Genese, Wirkungen und ihre Messung* (389-400). Weinheim: Beltz.
- Oser, F. (2013). "I know how to do it, but I can't do it": Modeling competence profiles for future teachers and trainers. In S. Blömeke, O. Zlatkin-Troitschanskaia, C. Kuhn & J. Fege (Eds.), *Modeling and measuring competencies in higher education* (45-59). Rotterdam: Sense. Zugriff am 17.8.2017. Verfügbar unter: https://link.springer.com/chapter/10.1007/978-94-6091-867-4_4
- Oser, F., Heinzer, S. & Salzmann, P. (2010). Die Messung der Qualität von professionellen Kompetenzprofilen von Lehrpersonen mit Hilfe der Einschätzung von Filmvignetten: Chancen und Grenzen des advokatorischen Ansatzes. *Unterrichtswissenschaft*, 38 (1), 5-28.
- Oser, F. K. & Baeriswyl, F. J. (2001). Choreographies of teaching: Bridging instruction to learning. In V. Richardson (Ed.), *Handbook of research on teaching* (1031-1065). Washington, DC: American Educational Research Association.
- Oser, F. & Oelkers, J. (2001). Einleitung. In F. Oser & J. Oelkers (Hrsg.), *Die Wirksamkeit der Lehrerbildungssysteme. Von der Allrounderbildung zur Ausbildung professioneller Standards* (15-33). Chur: Rüegger.
- Ozogul, G., Olina, Z. & Sullivan, H. (2008a). Teacher, self and peer evaluation of lesson plans written by preservice teachers. *Educational Technology Research and Development*, 56 (2), 181-201.
- Park, S. & Chen, Y.-C. (2012). Mapping out the integration of the components of pedagogical content knowledge (PCK): Examples from high school biology classrooms. *Journal of Research in Science Teaching*, 49 (7), 922-941.

- Park, S., Jang, J.-Y., Chen, Y.-C. & Jung, J. (2011). Is pedagogical content knowledge (PCK) necessary for reformed science teaching?: Evidence from an empirical study. *Research in Science Education*, 41 (2), 245-260.
- Park, S. & Oliver, J. S. (2008). Revisiting the conceptualisation of pedagogical content knowledge (PCK): PCK as a conceptual tool to understand teachers as professionals. *Research in Science Education*, 38 (3), 261-284.
- Park, S. & Suh, J. K. (2015). From portraying toward assessing PCK: Drivers, dilemmas, and directions for future research. In A. Berry, P. Friedrichsen & J. Loughran (Eds.), *Re-examining pedagogical content knowledge in science education* (104-119). New York: Routledge.
- Pauli, C. & Reusser, K. (2003). Unterrichtsskripts im schweizerischen und im deutschen Mathematikunterricht. *Unterrichtswissenschaft*, 31 (3), 238-272.
- Pauli, C. & Reusser, K. (2006). Von international vergleichenden Video Surveys zur videobasierten Unterrichtsforschung und -entwicklung. *Zeitschrift für Pädagogik*, 52 (6), 774-798.
- Pauli, C. & Reusser, K. (2009). Zum Einfluss von Professionalität auf die Qualität von Lehr-Lern-Prozessen. In O. Zlatkin-Troitschanskaia, K. Beck, D. Sembill, R. Nickolaus & R. Mulder (Hrsg.), *Lehrerprofessionalität. Bedingungen, Genese, Wirkungen und ihre Messung* (679-689). Weinheim: Beltz.
- Porter, A. C. (2002). Measuring the content of instruction: Uses in research and practice. *Educational Researcher*, 31 (7), 3-14.
- Praetorius, A.-K., Pauli, C., Reusser, K., Rakoczy, K. & Klieme, E. (2014). One lesson is all you need? Stability of instructional quality across lessons. *Learning and Instruction*, 31 (1), 2-12.
- Prenzel, M. (2000). Steigerung der Effizienz des mathematisch-naturwissenschaftlichen Unterrichts: Ein Modellversuchsprogramm von Bund und Ländern. *Unterrichtswissenschaft*, 28 (2), 103-126.
- Prenzel, M., Artelt, C., Baumert, J., Blum, W., Hammann, M. & Klieme, E., Pekrun, R. (Hrsg.). (2007). *PISA 2006. Die Ergebnisse der dritten internationalen Vergleichsstudie*. Münster: Waxmann.
- Prenzel, M. & Seidel, T. (2008). Erwerb naturwissenschaftlicher Kompetenzen. In W. Schneider & M. Hasselhorn (Hrsg.), *Handbuch Entwicklungspsychologie / Pädagogische Psychologie* (608-618). Göttingen: Hogrefe.

- Prenzel, M., Seidel, T., Lehrke, M., Rimmmler, R., Duit, R., Euler, M. et al. (2002). Lehr-Lernprozesse im Physikunterricht – eine Videostudie. In M. Prenzel & J. Doll (Hrsg.), *Bildungsqualität von Schule: Schulische und außerschulische Bedingungen mathematischer, naturwissenschaftlicher und überfachlicher Kompetenzen* (139-156). Weinheim: Beltz.
- Priemer, B. (2011). Was ist das Offene beim offenen Experimentieren?, *Zeitschrift für Didaktik der Naturwissenschaften*, 17, 315-337.
- Rakoczy, K. & Pauli, C. (2006). Hoch inferentes Rating: Beurteilung der Qualität unterrichtlicher Prozesse. In I. Hugener, C. Pauli & K. Reusser (Hrsg.), *Dokumentation der Erhebungs- und Auswertungsinstrumente zur schweizerisch-deutschen Video-studie „Unterrichtsqualität, Lernverhalten und mathematisches Verständnis“*. 3. Videoanalyse (206-233). Frankfurt a. Main: GEPP u.a.
- Ramseier, E., Labudde, P. & Adamina, M. (2011). Validierung des Kompetenzmodells HarmoS Naturwissenschaften: Fazite und Defizite. *Zeitschrift für Didaktik der Naturwissenschaften*, 17, 7-33.
- Rehm, M. & Bölsterli, K. (2014). Entwicklung von Unterrichtsvignetten. In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *Methoden in der naturwissenschaftsdidaktischen Forschung* (213-225). Berlin, Heidelberg: Springer.
- Reinisch, H. (2009). Zur Professionalität als Expertise. In O. Zlatkin-Troitschanskaia, K. Beck, D. Sembill, R. Nickolaus & R. Mulder (Hrsg.), *Lehrprofessionalität. Bedingungen, Genese, Wirkungen und ihre Messung* (33-43). Weinheim: Beltz.
- Reusser, K. & Halbheer, U. (2008). Bildungsstandards als Ausgangspunkt für Unterrichtsentwicklung. *Beiträge zur Lehrerbildung*, 26 (3), 304-317.
- Reusser, K. (2008). Empirisch fundierte Didaktik — didaktisch fundierte Unterrichtsforschung. In M.A. Meyer, M. Prenzel & S. Hellekamps (Hrsg.), *Perspektiven der Didaktik: Zeitschrift für Erziehungswissenschaft* (219-237). Wiesbaden: VS.
- Reusser, K. (2014). Kompetenzorientierung als Leitbegriff der Didaktik. *Beiträge zur Lehrerinnen- und Lehrerbildung*, 32 (3), 325-339.
- Reusser, K. & Pauli, C. (2010). Unterrichtsgestaltung und Unterrichtsqualität – Ergebnisse einer internationalen und schweizerischen Videostudie zum Mathematikunterricht: Einleitung und Überblick. In K. Reusser, C. Pauli & M. Waldis (Hrsg.), *Unterrichtsgestaltung und Unterrichtsqualität. Ergebnisse einer internationalen und schweizerischen Videostudie zum Mathematikunterricht* (S. 9-32). Münster: Waxmann.

- Riegel, U. (2013). Videobasierte Kompetenzforschung in den Fachdidaktiken. Einleitung. In U. Riegel & K. Macha (Hrsg.), *Videobasierte Kompetenzforschung in den Fachdidaktiken* (79-96). Münster: Waxmann.
- Riegel, U. & Macha, K. (2013). *Videobasierte Kompetenzforschung in den Fachdidaktiken*. Münster: Waxmann.
- Riese, J. (2009). Professionelles Wissen und professionelle Handlungskompetenz von (angehenden) Physiklehrkräften. Berlin: Logos.
- Riese, J. & Reinhold, P. (2010). Empirische Erkenntnisse zur Struktur professioneller Handlungskompetenz von angehenden Physiklehrkräften. *Zeitschrift für Didaktik der Naturwissenschaften*, (16).
- Riese, J. & Reinhold, P. (2012a). Kompetenzen von Lehramtstudierenden in Physik (Fachdidaktische Forschungen). In H. Bayrhuber, U. Harms, B. Muszynski, B. Ralle, M. Rothnagel, L.-H. Schön et al. (Hrsg.), *Formate Fachdidaktischer Forschung. Empirische Projekte – historische Analysen – theoretische Grundlegungen. Fachdidaktische Forschungen* (Bd. 2, 297-314). Münster: Waxmann.
- Riese, J. & Reinhold, P. (2012b). Die professionelle Kompetenz angehender Physiklehrkräfte in verschiedenen Ausbildungsformen. Empirische Hinweise für eine Verbesserung des Lehramtsstudiums. *Zeitschrift für Erziehungswissenschaft*, 15, 111-143.
- Roberts, D. A. (2007). Scientific literacy/Science literacy. In S. K. Abell & N. G. Lederman (Eds.), *Handbook of Research on Science Education* (S. 729-780).
- Roberts, R. & Gott, R. (1999). Procedural understanding: its place in the biology curriculum. *School Science Review*, 81 (294), 19-25.
- Roberts, R. & Gott, R. (2003). Assessment of biology investigations. *Journal of Biological Education*, 37 (3), 114-121.
- Roehrig, G. H. & Luft, J. A. (2004). Research Report. *International Journal of Science Education*, 26 (1), 3-24.
- Roick, T. & Henschel, S. (2015). Strategie zur Validierung von Kompetenzstrukturmodellen. In U. Riegel, S. Schubert, G. Siebert-Ott & K. Macha (Hrsg.), *Kompetenzmodellierung und Kompetenzmessung in den Fachdidaktiken* (S. 11-28). Münster: Waxmann.

- Rönnebeck, S., Bernholt, S. & Ropohl, M. (2016). Searching for a common ground – A literature review of empirical research on scientific inquiry activities. *Studies in Science Education*, 52 (2), 161-197.
- Roters, B. (2012). *Professionalisierung durch Reflexion in der Lehrerbildung. Eine empirische Studie an einer deutschen und US-amerikanischen Universität*. Münster: Waxmann.
- Roth, K. J., Druker, S. L., Garnier, H. E., Lemmens, M., Chen, C., Kawanaka, T. et al. (2006). *Teaching Science in five countries: Results from the TIMSS 1999 Video Study. Statistical Analysis Report. NCES 2006-011*. ED Pubs, P. Zugriff am 19.8.2017. Verfügbar unter: <https://eric.ed.gov/?id=ED491193>
- Ruiz-Primo, M. A. & Shavelson, R. J. (1996). Rhetoric and reality in science performance assessments: An update. *Journal of Research in Science Teaching*, 33 (10), 1045-1063.
- Ruys, I., Keer, H. V. & Aelterman, A. (2012). Examining pre-service teacher competence in lesson planning pertaining to collaborative learning. *Journal of Curriculum Studies*, 44 (3), 349-379.
- Ryle, G. (2009). *The concept of mind – 60 anniversary edition*. London and New York: Routledge Taylor and Francis Group.
- Sauter, W. & Staudt, A.-K. (2016). Methoden der Kompetenzmessung. In W. Sauter & A.-K. Staudt (Hrsg.) *Kompetenzmessung in der Praxis* (7-21), essentials. Wiesbaden: Springer.
- Schank, R. & Abelson, R. (1977). *Scripts, plans, goals and understanding*. Hillsdale, NJ: Erlbaum.
- Schaper, N. (2009). Aufgabenfelder und Perspektiven bei der Kompetenzmodellierung und -messung in der Lehrerbildung. *Lehrerbildung auf dem Prüfstand 2009* (166-199). Landau: Empirische Pädagogik.
- Schaper, N. (2014). Validitätsaspekte von Kompetenzmodellen und -tests für hochschulische Kompetenzdomänen. In F. Musekamp & G. Spöttl (Hrsg.), *Kompetenz im Studium und in der Arbeitswelt. Nationale und internationale Ansätze zur Erfassung von Ingenieurkompetenzen* (21-48). Frankfurt, M.: Lang.
- Schaper, N., Hilligus, A. H. & Reinhold, P. (2009). Kompetenzmodellierung und -messung in der Lehrerbildung. *Lehrerbildung auf dem Prüfstand 2009* (1-9). Landau: Empirische Pädagogik.

- Schecker, H. (2014). Überprüfung der Konsistenz von Itemgruppen mit Cronbachs alpha (chapter online-Zusatzmaterial). In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *Methoden in der naturwissenschaftsdidaktischen Forschung* (1-7). Berlin: Springer. Zugriff am 1.3.2019. Verfügbar unter <http://static.springer.com/sgw/documents/1426184/application/pdf/Cronbach+Alpha.pdf>
- Schecker, H. & Parchmann, I. (2006). Modellierung naturwissenschaftlicher Kompetenz. *Zeitschrift für Didaktik der Naturwissenschaften*, 12.
- Schendera, C. F. (2014). *Regressionsanalyse mit SPSS* (2. korrig. und aktual. Aufl.). Oldenburg: De Gruyter.
- Schermelleh-Engel, K. & Werner, C. S. (2012). Methoden der Reliabilitätsbestimmung. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (2., aktual. u. überarb. Aufl., 119-142). Berlin: Springer.
- Schmelzing, S. (2010). *Das fachdidaktische Wissen von Biologielehrkräften: Konzeptionalisierung, Diagnostik, Struktur und Entwicklung im Rahmen der Biologielehrerbildung*. Berlin: Logos.
- Schmelzing, S., van Driel, J. H., Jüttner, M., Brandenbusch, S., Sandmann, A. & Neuhaus, B. J. (2013). Development, evaluation, and validation of a paper-and-pencil test for measuring two components of biology teachers' pedagogical content knowledge concerning the „Cardiovascular System“. *International Journal of Science and Mathematics Education*, 11, 1369-1390.
- Schmelzing, S., Wüsten, S., Sandmann, A. & Neuhaus, B. (2008). Das fachdidaktische Wissen der Lehrkraft als Einflussfaktor für die Unterrichtsqualität im Biologieunterricht. *Erkenntnisweg Biologiedidaktik*, 7, 159-168.
- Schmelzing, S., Wüsten, S., Sandmann, A. & Neuhaus, B. (2010). Fachdidaktisches Wissen und Reflektieren im Querschnitt der Biologielehrerbildung. *Zeitschrift für Didaktik der Naturwissenschaften*, 16.
- Schmiemann, P. & Lücken, M. (2014). Validität – Misst mein Test, was er soll? In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *Methoden in der naturwissenschaftsdidaktischen Forschung* (107-118). Berlin: Springer.
- Schnebel, S., Kreis, A. & Musow, S. (2017). Wie schätzen Studierende ihre Planungskompetenz und den Nutzen kooperativer Planung ein? – Ergebnisse einer Interventionsstudie zu Peer Coaching in der Lehrpersonenausbildung nach dem Ansatz des Kollegialen Unterrichtacoachings. *Lehrerbildung auf dem Prüfstand* (Sonderheft, 107-128). Landau: Empirische Pädagogik.

- Schneider, R. M. & Plasman, K. (2011). Science teacher learning progressions: A review of science teachers' pedagogical content knowledge development. *Review of Educational Research*, 81 (4), 530-565.
- Schön, D. A. (1983). *The reflective practitioner. How professionals think in action*. New York: Basic books.
- Schreiber, N. (2012). Diagnostik experimenteller Kompetenz: Validierung technologiegestützter Testverfahren im Rahmen eines Kompetenzstrukturmodells. Berlin: Logos.
- Schüle, C., Besa, K.-S. & Arnold, K.-H. (2017). Entwicklung und Validierung eines Testinstruments zur Erfassung allgemeindidaktischer Kompetenz. In S. Wernke & K. Zierer (Hrsg.), *Die Unterrichtsplanung: Ein in Vergessenheit geratener Kompetenzbereich?! Status Quo und Perspektiven aus Sicht der empirischen Forschung* (17-31). Bad Heilbrunn: Julius Klinkhardt.
- Schulz, A. (2011). *Experimentierspezifische Qualitätsmerkmale im Chemieunterricht: Eine Videostudie*. Berlin: Logos.
- Schulz, A., Rieß, W., Wirtz, M. & Barzel, B. (2012). Experimentieren im mathematisch-naturwissenschaftlichen Unterricht: Schüler lernen wissenschaftlich denken und arbeiten (1. Aufl.). Münster: Waxmann.
- Schulz, W. (1980). *Unterrichtsplanung*. München: Urban & Schwarzenberg.
- Schwartz, R., Lederman, N. G. & Crawford, B. A. (2004). Developing views of nature of science in an authentic context: An explicit approach to bridging the gap between nature of science and scientific inquiry. *Science Teacher Education*, 610-645.
- Seeber, S. & Nickolaus, R. (2010). Kompetenzmessung in der beruflichen Bildung. *Berufsbildung in Wissenschaft und Praxis*, 1, 10-13.
- Seel, A. (1997). Von der Unterrichtsplanung zum konkreten Lehrerhandeln – Eine Untersuchung zum Zusammenhang von Planung und Durchführung von Unterricht bei Hauptschullehrerstudentinnen. *Unterrichtswissenschaft*, 25 (3).
- Seel, A. (2011). Lernwirksame Unterrichtsbedingungen in der Unterrichtsplanung berücksichtigen. *Jahrbuch für Allgemeine Didaktik*, 31-45.
- Seidel, T. (2014a). Lehrerhandeln im Unterricht. In E. Terhart, H. Bennewith & M. Rothland (Hrsg.), *Handbuch der Forschung zum Lehrerberuf* (2. Aufl., 781-806). Münster: Waxmann.

- Seidel, T. (2014b). Angebots-Nutzungs-Modelle in der Unterrichtspsychologie. *Zeitschrift für Pädagogik*, 6, 850-866.
- Seidel, T. & Prenzel, M. (2003). Videoanalyse als Methode in der Lehr- und Lernforschung. *Journal für Lehrerinnen- und Lehrerbildung*, 1, 54-61.
- Seidel, T. & Prenzel, M. (2004). Muster unterrichtlicher Aktivitäten im Physikunterricht. In J. Doll & M. Prenzel (Hrsg.), *Bildungsqualität von Schule: Lehrerprofessionalisierung, Unterrichtsentwicklung und Schülerförderung als Strategien der Qualitätsverbesserung* (177-194). Münster: Waxmann.
- Seidel, T., Prenzel, M., Duit, R. & Lehrke, M. (2003). Technischer Bericht zur Videostudie „Lehr-Lern-Prozesse im Physikunterricht“. Kiel: IPN.
- Seidel, T., Prenzel, M., Rimmele, R., Dalehefte, I. M., Herweg, C., Kobarg, M. et al. (2006). Blicke auf den Physikunterricht. Ergebnisse der IPN Videostudie. *Zeitschrift für Pädagogik*, 52 (6), 799-821.
- Seidel, T., Prenzel, M., Schwindt, K., Rimmele, R., Kobarg, M. & Dalehefte, I. M. (2009). The link between teaching and learning – Investigating effects of physics teaching on student learning in the context of the IPN Video Study. In T. Janik & T. Seidel (Eds.), *The power of video studies in investigating teaching and learning in the classroom* (161-180). Münster: Waxmann.
- Seidel, T. & Reiss, K. (2014). Lerngelegenheiten im Unterricht. *Pädagogische Psychologie* (253-276). Weinheim: Beltz.
- Seidel, T., Rimmele, R. & Prenzel, M. (2005). Clarity and coherence of lesson goals as a scaffold for student learning. *Learning and Instruction*, 15 (6), 539-556.
- Seidel, T. & Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: The role of theory and research design in disentangling meta-analysis results. *Review of Educational Research*, 77 (4), 454-499.
- Seifried, J. & Ziegler, B. (2009). Domänenspezifische Professionalität. In O. Zlatkin-Troitschanskaia, K. Beck, D. Sembill, R. Nickolaus & R. Mulder (Hrsg.), *Lehrprofessionalität. Bedingungen, Genese, Wirkungen und ihre Messung* (83-92). Weinheim: Beltz.
- Shavelson, R. (1983). Review of research on teachers' pedagogical judgments, plans, and decisions. *The Elementary School Journal*, 392-413.
- Shavelson, R. J. (2010). On the measurement of competency. *Empirical Research in Vocational Education and Training*, 2 (1), 41-63.

- Shavelson, R. J. (2013). Modeling competence profiles for future teachers and trainers. In S. Blömeke, O. Zlatkin-Troitschanskaia, C. Kuhn & J. Fege (Eds.), *Modeling and measuring competencies in higher education: Tasks and challenges* (29-43). Rotterdam: Sense.
- Shavelson, R. J., Ruiz-Primo, M. A. & Wiley, E. W. (2005). Windows into the mind. *Higher Education*, 49 (4), 413-430.
- Shavelson, R. J. & Stern, P. (1981). Research on teachers' pedagogical thoughts, judgments, decisions, and behavior. *Review of Educational Research*, 455-498.
- Shavelson, R. J. & Webb, N. M. (2006). Generalizability theory. In J. L. Green, G. Camilli, P. B. Elmore, A. Skukauskaiti & E. Grace (Eds.), *Handbook of complementary methods in education research* (599-612). United States of America: American Educational Research Association.
- Shulman, L. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, 57 (1), 1-23.
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15 (2), 4-14.
- Shulman, L. S. (2015a). PCK. Its genesis and exodus. In A. Berry, P. Friedrichsen & J. Loughran (Eds.), *Re-examining pedagogical content knowledge in science education* (3-13). New York: Routledge.
- Sim, J. & Wright, C. C. (2005). The kappa statistic in reliability studies: Use, interpretation, and sample size requirements. *Physical Therapy*, 85 (3), 257-268.
- Smit, R., Rietz, F. & Kreis, A. (2017). What are the effects of science lesson planning in peers? – Analysis of attitudes and knowledge based on an actor – partner interdependence model. *Research in Science Education*, 1-18.
- Smit, R., Weitzel, H., Blank, R., Rietz, F., Tardent, J. & Robin, N. (2017). Interplay of secondary pre-service teacher content knowledge (CK), pedagogical content knowledge (PCK) and attitudes regarding scientific inquiry teaching within teacher training. *Research in Science & Technological Education*, 35 (4), 477-499.
- Smith, P. S. & Banilower, E. R. (2015). Assessing PCK. A new application on the uncertain principle. In A. Berry, P. Friedrichsen & J. Loughran (Eds.), *Re-examining pedagogical content knowledge in science education* (88-103). New York and London: Routledge.

- So, W. W. (1997). A study of teacher cognition in planning elementary science lessons. *Research in Science Education*, 27 (1), 71-86.
- Spooner, F., Baker, J. N., Harris, A. A., Ahlgrim-Delzell, L. & Browder, D. M. (2007). Effects of training in universal design for learning on lesson plan development, *Remedial and Special Education*, 28 (2), 108-116.
- Squire, L. R., Stark, C. E. L. & Clark, R. E. (2004). The medial temporal lobe. *Annual Review of Neuroscience*, 27, 279-306.
- Standop, J. & Jürgens, E. (2015). *Unterricht planen, gestalten und evaluieren*. Bad Heilbrunn: Klinkhardt.
- Staub, F. & Kreis, A. (2013). Fachspezifisches Unterrichtscoaching in der Aus- und Weiterbildung von Lehrpersonen. *Journal für Lehrerbildung*, 2, 8-13.
- Stender, A. (2014). *Unterrichtsplanung: Vom Wissen zum Handeln: Theoretische Entwicklung und empirische Überprüfung des Transformationsmodells der Unterrichtsplanung*. Berlin: Logos.
- Stender, A., Brückmann, M. & Neumann, K. (2015). Vom Professionswissen zum kompetenten Handeln im Unterricht: Rolle der Unterrichtsplanung. *Beiträge zur Lehrerinnen- und Lehrerbildung*, 33 (1), 121-133.
- Stender, A., Brückmann, M. & Neumann, K. (2017). Transformation of topic-specific professional knowledge into personal pedagogical content knowledge through lesson planning. *International Journal of Science Education*, 39 (12), 1690-1714.
- Stender, A., Geller, C., Neumann, K. & Fischer, H. E. (2013). Der Einfluss der Unterrichtstaktung auf die Strukturiertheit und Abgeschlossenheit von Lernprozessen. *Zeitschrift für Didaktik der Naturwissenschaften*, 19, 189-207.
- Stern, E., Zeyer, A. & Metzger, S. (2009). *Expertise zu Naturwissenschaft und Technik in der Allgemeinbildung im Kanton Zürich*. Zürich: Zürcher Hochschule für Schulpädagogik und Fachdidaktik. Zugriff am 31.1.2017. Verfügbar unter: www.bi.zh.ch/dam/bildungsdirektion/direktion/bildungsplanung/projekte/natech/ZHSF_Expertise_NATECH.pdf
- Stigler, J. W., Gallimore, R. & Hiebert, J. (2000). Using video surveys to compare classrooms and teaching across cultures: Examples and lessons from the TIMSS video studies. *Educational Psychologist*, 35 (2), 87-100.

- Stigler, J. W., Gonzalez, P., Kawanaka, T., Knoll, S. & Serrano, A. (1999). *The TIMSS videotape classroom study. Methods and findings from an exploratory research project on eighth-grade mathematics instruction in Germany, Japan and the United States*. Washington D. C.: U.S. Department of Education.
- Stiller, J., Straube, P., Matthesius, S., Hartmann, S., Tiemann, R., Nordmeier, V. et al. (2014). Fachmethodische Kompetenzen im Lehramtstudium – das Projekt KoWADIS. In S. Bernholt (Hrsg.), *Naturwissenschaftliche Bildung zwischen Science- und Fachunterricht. Gesellschaft für Didaktik der Chemie und Physik, Jahrestagung in München 2013* (156-158). Kiel: IPN.
- Straka, G. A. & Macke, G. (1979). Klafkis Ansatz einer kritisch-konstruktiven Didaktik und das Konzept einer Lehr-Lern-Theorie - Alternative oder Ergänzung? *Die Deutsche Schule*, 71 (2), 111-113.
- Straka, G. A. & Macke, G. (2009). *Neue Einsichten in Lehren, Lernen und Kompetenz*. ITB-Forschungsbericht. Bremen: Institut Technik und Bildung (ITB), Universität Bremen.
- Stronge, J. H. (2006). Teacher evaluation and school improvement: Improving the educational landscape. In J. H. Stronge (Ed.), *Evaluating teaching: A guide to current thinking and best practice* (2. Aufl., 1-23). Thousand Oaks, CA: Corwin.
- Sweller, J., van Merriënboer, J. J. G. & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review*, 1-32.
- Tamir, P. (1988). Subject matter and related pedagogical knowledge in teacher education. *Teacher and Teacher Education*, 4 (2), 99-110.
- Tenorth, H.-E. (2006). Professionalität im Lehrerberuf. Ratlosigkeit der Theorie, gelingende Praxis. *Zeitschrift für Erziehungswissenschaft*, 9 (4), 580-597.
- Tenorth, H.-E. (2013). „Fachdidaktiken“ – ihre historische Entwicklung im Kontext pädagogischer Professionalisierung. In K.-P. Hufer & D. Richter (Hrsg.), *Politische Bildung als Profession Verständnisse und Forschungen. Perspektiven politischer Bildung* (20-32). Bonn: Bundeszentrale für politische Bildung.
- Tepner, O., Borowski, A., Dollny, S., Fischer, H. E., Jüttner, M., Kirschner, S. et al. (2012). Modell zur Entwicklung von Testitems zur Erfassung des Professionswissens von Lehrkräften in den Naturwissenschaften. *Zeitschrift für Didaktik der Naturwissenschaften*, 18, 7-28.

- Terbrügge, A. (2001). *Unterrichtsplanung zwischen didaktischen Ansprüchen und alltäglicher Berufs Anforderung. Eine empirische Studie zum Planungshandeln von Lehrerinnen und Lehrern in den Fächern Deutsch, Mathematik und Chemie*. Frankfurt am Main: Peter Lang.
- Terhart, E. (2000). *Perspektiven der Lehrerbildung in Deutschland. Abschlussbericht der von der Kultusministerkonferenz eingesetzten Kommission. Abschlussbericht*. Weinheim: Beltz.
- Terhart, E. (2002). *Standards für die Lehrerbildung. Eine Expertise für die Kultusministerkonferenz*. Münster: Universität Münster.
- Terhart, E. (2008). Allgemeine Didaktik: Traditionen, Neuanfänge, Herausforderungen. In M. A. Meyer, M. Prenzel & S. Hellekamps (Hrsg.), *Perspektiven der Didaktik* (13-34). Wiesbaden: VS.
- Terhart, E. (2011). Lehrerberuf und Professionalität. Gewandeltes Begriffsverständnis – neue Herausforderungen. In W. Helpster & R. Tippelt (Hrsg.), *Pädagogische Professionalität* (202-224). Weinheim: Beltz.
- Tesch, M. (2005). *Das Experiment im Physikunterricht. Didaktische Konzepte und Ergebnisse einer Videostudie*. Berlin: Logos.
- Tesch, M. & Duit, R. (2004). Experimentieren im Physikunterricht – Ergebnisse einer Videostudie. *Zeitschrift für Didaktik der Naturwissenschaften*, 10, 51-69.
- Tillema, H. H. (2009a). Assessment for learning to teach appraisal of practice teaching lessons by mentors, supervisors, and student teachers. *Journal of Teacher Education*, 60 (2), 155-167.
- Tillmann, K.-J. (2011). Konzepte der Forschung zum Lehrerberuf. In E. Terhart, H. Bennewitz & M. Rothland (Hrsg.), *Handbuch der Forschung zum Lehrerberuf* (232-240). Münster: Waxmann.
- Tulodziecki, G., Herzig, B. & Blömeke, S. (2017). *Gestaltung von Unterricht*. Bad Heilbrunn: Klinkhardt.
- Urban, D. & Mayerl, J. (2014). Datenqualität und Messmodelle. In D. Urban & J. Mayerl (Hrsg.), *Strukturgleichungsmodellierung: Ein Ratgeber für die Praxis* (117-158). Wiesbaden: Springer.
- van Dijk, E. M. & Kattmann, U. (2007). A research model for the study of science teachers' PCK and improving teacher education. *Teaching and Teacher Education*, 23 (6), 885-897.

- van Dijk, E. M. & Kattmann, U. (2010). Evolution im Unterricht: Eine Studie über fachdidaktisches Wissen von Lehrerinnen und Lehrern. *Zeitschrift für Didaktik der Naturwissenschaften*, 16.
- van der Valk, A. E. & Broekman, H. (1999). The lesson preparation method: A way of investigating pre-service teachers' pedagogical content knowledge. *European Journal of Teacher Education*, 22 (1), 11-22.
- van Driel, J. H., Verloop, N. & de Vos, W. (1998). Developing science teachers' pedagogical content knowledge. *Journal of Research in Science Teaching*, 35 (6), 673-695.
- Verloop, N., van Driel, J. & Meijer, P. (2001). Teacher knowledge and the knowledge base of teaching. *International Journal of Educational Research*, 35 (5), 441-461.
- Vogelsang, C. (2014). *Validierung eines Instruments zur Erfassung der professionellen Handlungskompetenz von (angehenden) Physiklehrkräften: Zusammenhangsanalysen zwischen Lehrerkompetenz und Lehrerperformanz*. Berlin: Logos.
- Vogelsang, C. & Reinhold, P. (2013). Zur Handlungsvalidität von Tests zum professionellen Wissen von Lehrkräften. *Zeitschrift für Didaktik der Naturwissenschaften*, 19, 103-128.
- Vogelsang, C. & Riese, J. (2017). Wann ist eine Unterrichtsplanung ‚gut‘? Planungsperformanz in Praxisratgebern zur Unterrichtsplanung. In S. Wernke & K. Zierer (Hrsg.), *Die Unterrichtsplanung: Ein in Vergessenheit geratener Kompetenzbereich?! Status Quo und Perspektiven aus Sicht der empirischen Forschung* (47-61). Bad Heilbrunn: Verlag Klinkhardt.
- Vorholzer, A., Aufschnaiter, C. v. & Kirschner, S. (2016). Entwicklung und Erprobung eines Tests zur Erfassung des Verständnisses experimenteller Denk- und Arbeitsweisen. *Zeitschrift für Didaktik der Naturwissenschaften*, 22 (1), 25-41.
- Voss, T. & Kunter, M. (2013). Teachers' general pedagogical/psychological knowledge. In J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Eds.), *Cognitive activation in the mathematics classroom and professional competence of teachers: Results from the COACTIV project* (207-228). New York: Springer.
- Voss, T., Kunter, M., Baumert, J. (2011). Assessing teacher candidates' general pedagogical/psychological knowledge: Test construction and validation. *Journal of Educational Psychology*, 103 (4), 952-969.

- Wadouh, J., Liu, N., Sandmann, A. & Neuhaus, B. J. (2014). The effect of knowledge linking levels in biology lessons upon students' knowledge structure. *International Journal of Science and Mathematics Education*, 12 (1), 25-47.
- Wadouh, J., Sandmann, A. & Neuhaus, B. (2009). Vernetzung im Biologieunterricht – deskriptive Befunde einer Videostudie. *Zeitschrift für Didaktik der Naturwissenschaften*, 15, 69-87.
- Wagner, S. (2014). *Handeln von Lehrpersonen beim naturwissenschaftlichen Lernen – Eine videobasierte Analyse des Unterstützungshandelns und seiner Bezüge zu Lehrervorstellungen*. Weingarten: Pädagogische Hochschule Weingarten.
- Wagner, W., Göllner, R., Werth, S., Voss, T., Schmitz, B. & Trautwein, U. (2016). Student and teacher ratings of instructional quality: Consistency of ratings over time, agreement, and predictive power. *Journal of Educational Psychology*, 108 (5), 705-721.
- Wahl, D. (2002). Mit Training vom trägen Wissen zum kompetenten Handeln? *Zeitschrift für Pädagogik*, 48 (2), 227-241.
- Wallace, C. S. & Kang, N.-H. (2004). An investigation of experienced secondary science teachers' beliefs about inquiry: An examination of competing belief sets. *Journal of Research in Science Teaching*, 41 (9), 936-960.
- Walpuski, M. & Schulz, A. (2011). Researching experimental phases in chemistry education using video analysis. In C. Bruguière, A. Tiberghien & P. Clément (Eds.), *Retrieved from http://www.esera.org/media/ebook/strand3/ebook-esera2011_WALPUSKI-03.pdf*. Gehalten auf der eBook Proceedings of the ESERA 2011 Conference: Part 3: Teaching and learning science.
- Warrens, M. J. (2012). Cohen's linearly weighted kappa is a weighted average. *Advances in Data Analysis and Classification*, 6 (1), 67-79.
- Weiber, R. & Mülhhaus, D. (2014). *Strukturgleichungsmodellierung. Eine anwendungsorientierte Einführung in die Kausalanalyse mit Hilfe von AMOS, SmartPLS und SPSS* (2. Aufl.). Berlin: Springer.
- Weinert, F. E. (1990). Unterrichtsexpertise – Ein Konzept zur Verringerung der Kluft zwischen zwei theoretischen Paradigmen. In L.-M. Alisch, J. Baumert & K. Beck (Hrsg.), *Professionswissen und Professionalisierung (Braunschweiger Studien zur Erziehungs- und Sozialarbeitswissenschaft)* (Bd. 28, 173-206). Braunschweig: Copy Center Colmsee.

- Weinert, F. E. (1996). „Der gute Lehrer“, „die gute Lehrerin“ im Spiegel der Wissenschaft. Was macht Lehrende wirksam und was führt zu ihrer Wirksamkeit? *Beiträge zur Lehrerbildung*, 14 (2), 141-151.
- Weinert, F. E. (1999). *Definition and selection of competencies concepts of competence*. München: Max Planck Institute for Psychological Research.
- Weinert, F. E. (2000). Lehren und Lernen für die Zukunft – Ansprüche an das Lernen in der Schule. *Pädagogische Nachrichten Rheinland-Pfalz*, 2, 1-16.
- Weinert, F. E. (2001a). Vergleichende Leistungsmessung in Schulen – eine umstrittene Selbstverständlichkeit. In F. E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (17-31). Weinheim: Beltz.
- Weinert, F. E. (2001b). Concept of competence: A conceptual clarification. In D. S. Rychen & L. H. Salganik (Eds.), *Defining and selecting key competencies* (45-65). Ashland, OH: Hogrefe & Huber.
- Weinert, F. E., Schrader, F.-W. & Helmke, A. (1990). Educational expertise: Closing the gap between educational research and classroom practice. *School Psychology International*, 11 (3), 163-180.
- Weingarten, J. & van Ackeren, I. (2017). Wie planen angehende Lehrkräfte ihren Unterricht? Empirische Befunde zur kompetenzorientierten Gestaltung von Lernangeboten. In S. Wernke & K. Zierer (Hrsg.), *Die Unterrichtsplanung: Ein in Vergessenheit geratener Kompetenzbereich?! Status Quo und Perspektiven aus Sicht der empirischen Forschung* (148-165). Bad Heilbrunn: Klinkhardt.
- Weitzel, H. (2015). Die Qualität von Unterrichtsplanungen zu hypothesengeleitetem Experimentieren analysieren. Vorstellung eines hoch inferenten Ratinginstruments und erster Ergebnisse. Gehalten auf dem SGBF Kongress 2015: Qualitäts- und Bildungsdiskurs, St. Gallen.
- Wellnitz, N. (2012). *Kompetenzstruktur und -niveaus von Methoden naturwissenschaftlicher Erkenntnisgewinnung*. Berlin: Logos.
- Wellnitz, N., Hecht, M., Heitmann, P., Kauertz, A., Mayer, J., Sumfleth, E. et al. (2017). Modellierung des Kompetenzteilbereichs naturwissenschaftliche Untersuchungen. *Zeitschrift für Erziehungswissenschaft*, 1-29.
- Wellnitz, N. & Mayer, J. (2013). Erkenntnismethoden in der Biologie – Entwicklung und Evaluation eines Kompetenzmodells. *Zeitschrift für Didaktik der Naturwissenschaften*, 19, 315-345.

- Werner, J., Wernke, S. & Zierer, K. (2017). Der Einfluss didaktischer Modelle auf die all-gemeindidaktische Unterrichtsplanungskompetenz von Lehramtsstudierenden. In S. Wernke & K. Zierer (Hrsg.), *Die Unterrichtsplanung: Ein in Vergessenheit geratener Kompetenzbereich?! Status Quo und Perspektiven aus Sicht der empirischen For-schung* (104-120). Bad Heilbrunn: Klinkhardt.
- Wernke, S. & Zierer, K. (2017). Die Unterrichtsplanung. Ein in Vergessenheit geratener Kompetenzbereich?! In S. Wernke & K. Zierer (Hrsg.), *Die Unterrichtsplanung: Ein in Vergessenheit geratener Kompetenzbereich?!: Status Quo und Perspektiven aus Sicht der empirischen Forschung T* (7-16). Bad Heilbrunn: Klinkhardt.
- Wiater, W. (2009). Lehrplan, Curriculum, Bildungsstandards. In K.-H. Arnold, U. Sand-fuchs & J. Wiechmann (Hrsg.), *Handbuch Unterricht* (2. aktual. Aufl., 140-143). Bad Heilbrunn: Klinkhardt.
- Widodo, A. & Duit, R. (2004). Konstruktivistische Sichtweisen vom Lehren und Lernen und die Praxis des Physikunterrichts. *Zeitschrift für Didaktik der Naturwissenschaften*, 10, 233-255.
- Wigfield, A. & Eccles, J. S. (2000). Expectancy-Value theory of achievement motivation. *Contemporary Educational Psychology*, 25 (1), 68-81.
- Wilhelm, M. (2007). Was ist guter Naturwissenschafts-Unterricht?, *Jg. 33* (Nr. 98), 67-87.
- Wilhelm, M. (2012). Kompetenzorientierten Unterricht konzipieren – am Beispiel der Na-turwissenschaften. *Haushalt in Bildung und Forschung*, 1 (3), 15-30.
- Wilhelm, M. & Kalcsics, K. (2017). *Lernwelten Natur – Mensch - Gesellschaft. Weiterbil-dung. Grundlagen und Planungsbeispiele, Praxisbuch 3. Zyklus*. Bern: Schulverlag Plus.
- Wilhelm, O. & Kunina-Habenicht, O. (2015). Pädagogisch psychologische Diagnostik. In E. Wild & J. Möller (Hrsg.), *Pädagogische Psychologie* (2. Aufl., 305-328). Berlin: Springer.
- Wilson, C. D., Taylor, J. A., Kowalski, S. M. & Carlson, J. (2010). The relative effects and equity of inquiry-based and commonplace science teaching on students' knowledge, reasoning, and argumentation. *Journal of Research in Science Teaching*, 47 (3), 276-301.

- Windschitl, M. (2003). Inquiry projects in science teacher education: What can investigative experiences reveal about teacher thinking and eventual classroom practice? *Science Education*, 87 (1), 112-143.
- Windschitl, M. (2004). Folk theories of "inquiry": How preservice teachers reproduce the discourse and practices of a theoretical scientific method. *Journal of Research in Science Teaching*, 41 (5), 481-512.
- Windschitl, M., Thompson, J. & Braaten, M. (2008). Beyond the scientific method: Model-based inquiry as a new paradigm of preference for school science investigations. *Science Education*, 92 (5), 941-967.
- Wirtz, M. & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität*. Göttingen: Hogrefe.
- Wüsten, S. (2010). *Allgemeine und fachspezifische Merkmale der Unterrichtsqualität im Fach Biologie: Eine Video- und Interventionsstudie*. Berlin: Logos.
- Wüsten, S., Schmelzing, S. & Sandmann, A. (2010). Sachstrukturdiagramme – Eine Methode zur Erfassung inhaltsspezifischer Merkmale der Unterrichtsqualität im Biologieunterricht. *Zeitschrift für Didaktik der Naturwissenschaften*, 16, 23-39.
- Wyss, C. (2013). *Unterricht und Reflexion. Eine mehrperspektivische Untersuchung der Unterrichts- und Reflexionskompetenz von Lehrkräften* (Empirische Wissenschaften). Münster: Waxmann.
- Yoon, H.-G., Joung, Y. J. & Kim, M. (2012). The challenges of science inquiry teaching for pre-service teachers in elementary classrooms: Difficulties on and under the scene. *Research in Science Education*, 42 (3), 589-608.
- Zierer, K. (2012). *Studien zur Allgemeinen Didaktik*. Baltmannsweiler: Schneider Hohengehren.
- Zierer, K. (2015). Educational expertise: the concept of 'mind frames' as an integrative model for professionalisation in teaching. *Oxford Review of Education*, 41 (6), 782-798.
- Zierer, K., Werner, J. & Wernke, S. (2015). Besser planen? Mit Modell! Empirisch basierte Überlegungen zur Entwicklung eines Planungskompetenzmodells. *DDS – Die Deutsche Schule*, 107 (4), 375-395.
- Zierer, K., Wernke, S. & Werner, J. (2015). Heimann, Schulz oder Klafki? *Zeitschrift für Pädagogik*, 61 (3), 429-451.

- Zlatkin-Troitschanskaia, O., Beck, K., Sembill, D., Nickolaus R. & Mulder, R. (2009a). Perspektiven auf „Lehrprofessionalität“. In O. Zlatkin-Troitschanskaia, K. Beck, D. Sembill, R. Nickolaus & R. Mulder (Hrsg.), *Lehrprofessionalität. Bedingungen, Genese, Wirkungen und ihre Messung* (13-32). Weinheim: Beltz.
- Zlatkin-Troitschanskaia, O., Beck, K., Sembill, D., Nickolaus, R. & Mulder, R. (Hrsg.). (2009b). *Lehrerprofessionalität. Bedingungen, Genese, Wirkungen und ihre Messung*. Weinheim: Beltz.

Dank

Es ist mir ein besonderes Anliegen, an dieser Stelle allen Menschen ganz herzlich zu danken, die mich im Rahmen dieses Dissertationsprojekts in verschiedenster Art und Weise unterstützt, begleitet und zum Gelingen dieser Arbeit beigetragen haben.

Insbesondere danke ich Markus Wilhelm für seine unterstützende, wertschätzende und von Gelassenheit geprägte Betreuung. Aufgrund seiner wertvollen Rückmeldungen und der damit verbundenen wissenschaftliche Förderung konnte ich inhaltlich und fachmethodisch viel profitieren.

Ein weiterer Dank geht an Markus Rehm für die zweite Betreuung und Begutachtung meiner Dissertation. Er hat mir Türen an der Pädagogischen Hochschule Heidelberg geöffnet und mich ebenfalls in einer sehr unkomplizierten, wertschätzenden und fruchtbaren Art unterstützt.

Danken möchte ich aber auch meinen Vorgesetzten der Pädagogischen Hochschule Zürich, die mir das Vertrauen geschenkt und mich bei der Realisierung dieser Dissertation unterstützt haben. Es sind dies Claude Walther, Esther Kamm, Christine Bieri, Hanspeter Pfirter, Luigi Bazzigher, Silja Rüedi und Andrea Widmer.

Ich danke auch Annelies Kreis und Stefanie Schnebel, die das Projekt KUBeX geleitet und mir die Möglichkeit eröffnet haben, mich als Fachdidaktikerin Biologie in diesem Forschungsprojekt einzubringen. Corinne Wyss danke ich für die im Rahmen des Projekts KubeX erfolgte spannende Zusammenarbeit. Danken möchte ich auch allen, im Projekt KUBeX tätigen Biologiedidaktikern, für den anregenden Fachaustausch und die Zusammenarbeit, insbesondere Holger Weizel, Florian Rietz, Robbert Smit und Robert Blank. Ein besonderer Dank geht weiter an Christine Marx, Wassilis Kassis und Urs Grob, die mich fachmethodisch sehr gut beraten haben.

Einen herzlichen Dank möchte ich zudem allen meinen Kolleginnen und Kollegen der Pädagogischen Hochschule Zürich aussprechen, die durch ihre vielfältige Unterstützung auch zum Gelingen dieser Arbeit beigetragen haben. Besonders erwähnen möchte ich Christoph Gut, dem ich für die fruchtbaren wissenschaftlichen Diskussionen, die wertvollen Rückmeldungen und die inspirierende Zusammenarbeit ganz herzlich danke. Ein besonderer Dank gilt auch Matthias Baer, der mich für dieses Vorhaben motiviert und mich stets mit wertvollen Rückmeldungen begleitet hat. Für weitere hilfreiche Kommentare zur Arbeit danke ich Maja Brückmann und Pitt Hild. Danken möchte ich aber auch meiner Freundin und Arbeitskollegin Beatrice Bürgler, die mich in dieser langen Zeit stets tatkräftig und mit hilfreichen Hinweisen unterstützte. Ein besonderer Dank geht auch an meine Schulfreundin

Claudia Schmid für das kritische Gegengelesen meiner Arbeit und die enorm hilfreichen und fruchtbaren Diskussionen. Der Aebli-Näf-Stiftung danke ich herzlich für die grosszügige Unterstützung im Rahmen eines Kurzaufenthalts (*short visit*) in Heidelberg, den ich für die Auswertung der Daten und das Verfassen erster Kapitel nutzen konnte

Ein inniger Dank geht auch an meine Eltern, insbesondere an meinen Vater, der mich für die Biologie, aber auch für die Wissenschaft begeistern konnte und für mich fachlich wie auch menschlich immer ein Vorbild war und sein wird. Den innigsten Dank möchte ich aber meinem Mann Marcel Kuster-Tardent für seine stete Unterstützung und sein Vertrauen aussprechen, die viel zum Gelingen dieser Arbeit beigetragen haben.